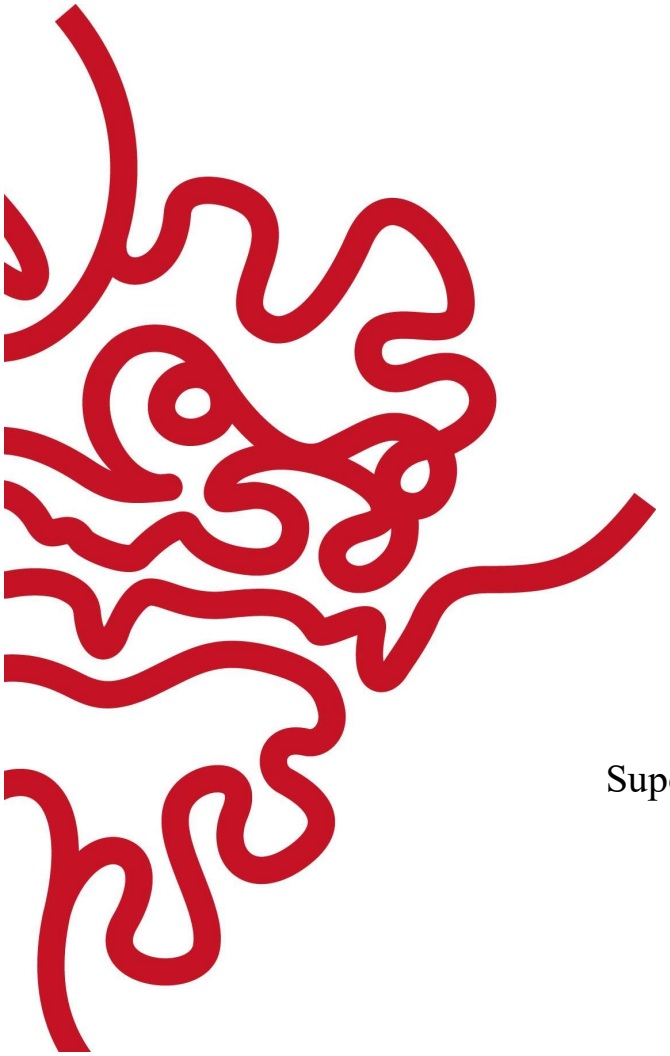
Investigation of Circular RNA Regulation by Cis and Trans Elements in Caenorhabditis Elegans

OKINAWA INSTITUTE OF SCIENCE AND TECHNOLOGY
GRADUATE UNIVERSITY

Thesis submitted for the degree
Doctor of Philosophy

Investigation of circular RNA regulation by *cis* and *trans* elements in *Caenorhabditis elegans*



by

Dong Cao

Supervisor: **Ichiro Maruyama**

2021-09

Declaration of Original and Sole Authorship

I, Dong Cao, declare that this thesis entitled "Investigation of circular RNA regulation by *cis* and *trans* elements in *Caenorhabditis elegans*" and the data presented in it are original and my own work.

I confirm that:

- No part of this work has previously been submitted for a degree at this or any other university.

- References to the work of others have been clearly acknowledged. Quotations from the work of others have been clearly indicated, and attributed to them.

- In cases where others have contributed to part of this work, such contribution has been clearly acknowledged and distinguished from my own work.

- None of this work has been previously published elsewhere, with the exception of the following:

  1. Cao, D. Reverse complementary matches simultaneously promote both back-splicing and exon-skipping. *BMC Genomics* **22**, 586, doi:10.1186/s12864-021-07910-w (2021).

  2. Cao, D. An autoregulation loop in *fust-1* for circular RNA regulation in *Caenorhabditis elegans*. *Genetics*, doi:10.1093/genetics/iyab145 (2021).

Date: 2021/09/15

Signature: 

Abstract

Circular RNAs (circRNAs) are regulatory molecules that show diverse functions. However, the regulation of circRNA formation is not yet well-understood. Through large-scale neuron isolation from the first larval stage of *Caenorhabditis elegans* followed by RNA sequencing with ribosomal RNA depletion, the first neuronal circRNA profile in *C. elegans* was obtained. Using circRNAs identified in this dataset, I performed an *in vivo* investigation of circRNA regulation by *cis* and *trans* elements.

Several neuronal circRNAs were knocked out by deleting one of the reverse complementary match (RCM) sequences flanking circRNA exon(s) (*cis* elements). Further, RCMs not only vigorously promote circRNA formation but also are beneficial for the skipping of exon(s) to be circularized. Through *in vivo* one-by-one mutagenesis of all the splicing sites and branch points required for exon-skipping and back-splicing in the *zip-2* gene, I showed that exon-skipping is not absolutely required for back-splicing, neither the other way. Instead, the coupled exon-skipping and back-splicing are promoted by RCMs directly at the same time.

As for *trans* elements that regulate circRNA in *C. elegans*, thirteen RNA binding proteins were screened, among which loss of FUST-1, the homolog of FUS, causes substantial downregulation of multiple circRNAs. Further, FUST-1 regulates circRNAs without affecting their cognate linear mRNA levels. When recognizing circRNA pre-mRNAs, FUST-1 can affect the coupled exon-skipping and circRNA formation in the same genes. In *zip-2*, the 5' splice sites for back splicing and exon skipping seem important for FUST-1's role in exon-skipping and back-splicing regulation, respectively. Two mutations (R446S and P447L) were introduced in FUST-1 to mimic the amyotrophic lateral sclerosis-related natural mutations in the nuclear localization signal of FUS (R524S and P525L). Both mutations dramatically affect circRNA levels. Moreover, I identified an autoregulation loop important for circRNA regulation in *fust-1*, where FUST-1, isoform a promotes the skipping of exon 5 of its own pre-mRNA, which produces FUST-1, isoform b with different N-terminal sequences. FUST-1, isoform a is the functional isoform in circRNA regulation. Although FUST-1, isoform b has the same functional domains as isoform a, it cannot regulate either exon-skipping or circRNA formation.

This thesis explored circRNA regulation *in vivo* using *C. elegans* as the model organism, providing new insights into mechanisms governing the relationship between back-splicing. The combinatorial regulation of circRNA by *cis* and *trans* elements supports a model of circRNA formation, where RCM sequences (*cis* elements) determine whether circRNA can be formed or not, and RBPs (*trans* elements) regulate how efficiently they can be produced.

Acknowledgments

My gratitude goes to everyone who helped me in these years of my Ph.D. study.

I would like to thank Prof. Ichiro Maruyama, my supervisor. Maruyama-sensei supported almost everything I asked for my Ph.D. work, giving me great freedom to explore whatever I wanted. Maruyama-sensei has high standards in doing science, which push me to think deeper when dealing with scientific problems.

I would like to thank Prof. Julia Khusnutdinova, Prof. Ichiro Masai, and Prof. Hidetoshi Saze for giving invaluable suggestions in both science and life.

I am very thankful:

… to Prof. Pedro Miura of the University of Nevada, Reno, who was the external examiner of my thesis proposal and is a great friend.

… to members I worked with in Maruyama Unit at OIST. Special thanks go to Dr. Takashi Murayama for training me on the techniques used in *C. elegans* field and the discussions we had. I also thank Miss Hitomi Ohtaki for the help with administrative work.

… to Dr. Adam Noris of the Southern Methodist University and Dr. Alex Parker of the University of Montreal for unselfish sharing of their strains and plasmids.

… to the staff of the sequencing section of OIST for their great help on the RNA-seq of my samples.

… to all the friends here in Okinawa.

This thesis is dedicated to my family, especially to my wife ZHANG Xuejun and my son CAO Ran. I love you.

Abbreviations

Ab: antibody

ADAR1: adenosine deaminase acting on RNA-1

AGO2: Argonaute 2

ANOVA: analysis of variance

AS: alternative splicing

BP: branch point

BSA: bovine serum albumin

BSJ: back-spliced junction

C. elegans: *Caenorhabditis elegans*

cas9: CRISPR-associated protein 9

CDK2: cyclin-dependent kinase 2

cDNA: complementary DNA

ceRNA: competing endogenous RNA

cGAS: cyclic GMP-AMP synthase

CI: chemotaxis index

circRNA: circular RNA

ciRNA: circular intronic RNA

Co-IP: co-immunoprecipitation

CRISPR: clustered regularly interspaced short palindromic repeats

crRNA: CRISPR RNA

Ct: threshold cycle

DAPI: 4',6-diamidino-2-phenylindole

ddPCR: droplet digital PCR

dsRNA: double-stranded RNA

DTT: 1,4-dithiothreitol

EGFR: epidermal growth factor receptor

EIciRNA: exon-intron circRNA

EMT: epithelial-mesenchymal transition

eSC: embryonic stem cell

EtBr: ethidium bromide

FACS: fluorescence-activated cell sorting

FBS: fetal bovine serum

FSC: forward scatter

FUS: fused in sarcoma

GBM: glioblastoma

GFP: green fluorescent protein

GO: gene ontology

gRNA: guide RNA

HDV: hepatitis delta (δ) virus

HEK: human embryonic kidney

hESC: human embryonic stem cell

hnRNP: heterogeneous nuclear ribonucleoprotein

ICS: intonic complementary sequencse

IME: intron-mediated enhancement

IR: inverted repeat

IRES: internal ribosome entry site

KD: knockdown

KO: knockout

L1: first larval stage

L4: fourth larval stage

lncRNA: long non-coding RNA

LT-HSCs: long-term hematopoietic stem cells

m^{6}A: *N*^{6}-methyladenosine

MBL: muscleblind

miRNA: microRNA

mRFP: monomeric red fluorescent protein

mRNA: messenger RNA

ncRNA: non-coding RNA

NGM: nematode growth media

NLS: nuclear localization signal

NTC: no-template control

PAGE: polyacrylamide gel electrophoresis

PCA: principal component analysis

PCR: polymerase chain reaction

PI: propidium iodide

Pol II: polymerase II

pre-mRNA: precursor mRNA

PVDF: polyvinylidene fluoride

QKI: Quaking

qPCR: quantitative PCR

RBP: RNA-binding protein

RCM: reverse complementary match

RNA ISH: RNA in situ hybridization

RNAi: RNA interference

RNA-seq: RNA sequencing

RRM: RNA-recognition motif

RRM20: RNA-binding motif protein 20

RT-PCR: reverse transcription PCR

RT-qPCR: quantitative RT-PCR

sd: standard deviation

SDS: sodium dodecyl sulfate

SEM: standard error of the mean

SFPQ: splicing factor proline/glutamine rich

siRNA: short interfering RNA

snRNP: small nuclear ribonucleoprotein

ss: splice site

ssODN: single-stranded oligo DNA

ssRNA: single-stranded RNA

TER: transcription elongation rate

TPM: transcripts per million reads

tracrRNA: trans-activating crRNA

UTR: untranslated region

ZnF: zinc-finger

Table of Content

List of Figures

List of Tables

1. Introduction

For most eukaryotes, nascent precursor messenger RNA (pre-mRNA) must undergo a splicing process, by which non-coding introns are removed and exons are joined together (Figure 1.1A). In many cases, splicing can generate a variety of mature mRNA isoforms, in which the composition of exons from the same pre-mRNA can be very different. This phenomenon is called alternative splicing (AS), and it is a fundamental mechanism to increase the diversity of the transcriptome. It is estimated that 92-94% of human genes are alternatively spliced (*1*). In general, transcript isoforms generated by AS are linear. However, recently, circular transcripts have been identified in various species, with most of them derived from coding genes (*2, 3*). In these circRNAs, the 5' and 3' ends are covalently joined to form a single-stranded circle. It is believed that most circRNAs are generated by "back-splicing," in which a splice donor downstream (5' splice site or 5'ss) is ligated to an upstream splice acceptor (3'ss) (Figure 1.1B and C). Introns can be retained in this process, forming exon-intron circRNAs (EIciRNAs) (*4*). In some cases, intron lariats can escape debranching and form circular intronic RNAs (ciRNAs) with 2',5'-phosphodiester bonds between the branch point nucleotides and the 5' splice sites (Figure 1.1D) (*5*).

Here, I summarize the discovery of circRNAs, the mechanisms that govern their biogenesis, and the functions of circRNAs.

1.1 History of circRNA discovery

The first RNA circles were reported in 1976 when Sanger et al. examined RNAs from four viroids and found that they were covalently closed single-stranded circular molecules (*6*). Subsequent sequencing of the viroid nucleic acid confirmed that these RNA molecules were indeed circRNAs (*7*). Later, in some fractions of yeast mitochondria RNA, Arnberg et al. reported the existence of circular molecules (*8*). circRNA molecules were later identified in self-splicing introns of ribosomal RNA precursors (*9, 10*) and hepatitis delta (δ) virus (HDV) (*11*). It was not until the 1990s that the first circRNA in human cells was identified. In 1991, Nigro et al. reported the existence of "scrambled exons" in the transcripts of a tumor suppressor gene called *deleted in colorectal carcinoma* (*12*). Exons in these transcripts were joined precisely at the consensus splice sites; however, the orders of the exons differ from those in their nascent transcripts (*12*). These "scrambled" transcripts were expressed at relatively low levels and were enriched in the non-polyadenylated fraction of RNA in the cytoplasm; hence, they were considered splicing errors (*12*). Subsequently, Cocquerelle et al. also reported "splicing with inverted order of exons" in the human *ETS1* gene (*13*). With further characterization, they showed that these transcripts corresponded to circRNAs with high stability (*14*). These transcripts that indicated the existence of circRNA were regarded as by-products or errors (mis-splicing) in canonical splicing steps due to their relatively low abundances (*12, 14*). However, some circRNAs were found to be abundant. The *Sry* gene in mouse testis produced a single-exon circRNA that represented more than 90% of the *Sry* transcripts (*15*). The circRNA derived from exon 2 of *NCX1* showed relatively high abundance across different species (*16*). The circRNA derived from exons in the mouse *formin* (*Fmn*) gene also presented very high levels (*17*). This evidence demonstrated the existence of circRNAs in different species, but due to the lack of high-throughput technologies, it was not possible to identify circRNAs systematically. Hence, during the subsequent several years, only a handful of circRNAs were reported (*18-20*). Thanks to the

development of RNA sequencing (RNA-seq) technology and bioinformatics algorithms, thousands of circRNAs were identified in an attempt to identify transcripts in cancer cells resulting from chromosomal rearrangements (*21*). These circRNAs were also identified in normal cells, indicating that they were not specific to cancer cells. This finding spurred the field of circRNA research, and in ensuing years, hundreds of thousands of circRNAs have been predicted in human, mouse, fly, worm, etc (*22*).

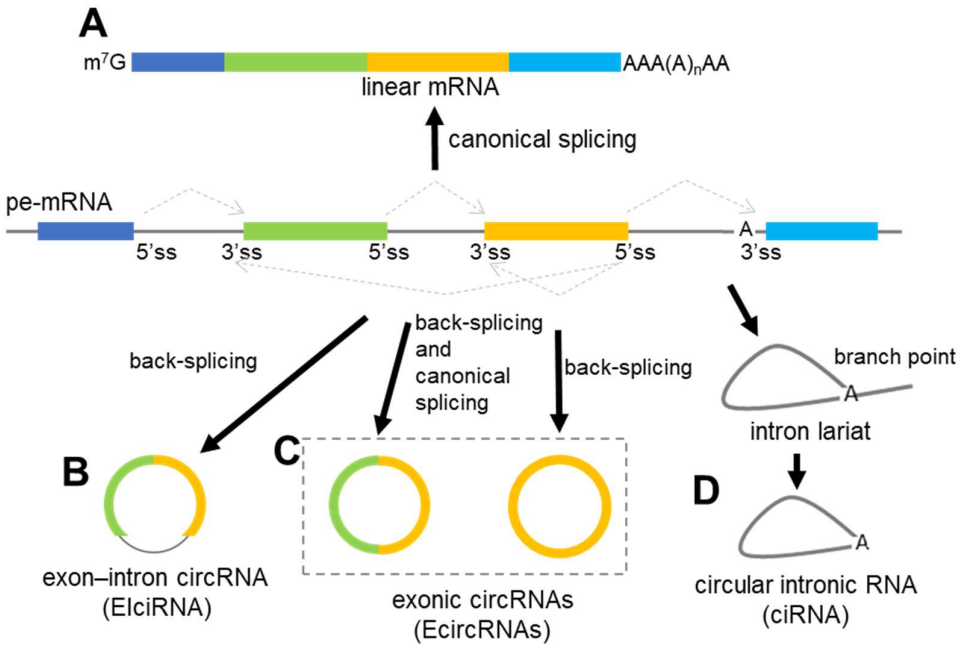


Figure 1.1 Types of circRNAs generated during splicing. (A) Canonically, a pre-mRNA molecule is spliced to form a mature linear mRNA with exons by joining the upstream 5'ss to downstream 3'ss. (B, C) Through back-splicing, circRNAs are formed. Introns can be retained in circRNAs, forming exon-intron circRNAs (EIciRNA). circRNAs formed by exons only are called exonic circRNAs. (D) Circular intronic RNAs (ciRNAs) are formed if intron lariats are trimmed and escaped from degradation.

1.2 circRNA Biogenesis

Although the detailed mechanism of circRNA biogenesis is still not fully understood, some progress has been achieved regarding elements involved in their production. Recent findings suggest that both *cis* and *trans* elements can regulate circRNA biogenesis.

1.2.1 Canonical splicing and back-splicing

The precise joining of canonical splice sites in reverse order in circRNAs suggests that spliceosomes are involved in circRNA formation (*12-14*). Indeed, exon(s) in pre-mRNAs can be circularized *in vitro* using nuclear extract, although in low efficiency (*23-27*). Several recent studies also showed that mutations of canonical splicing sites abolish circRNA formation (*28-31*). Especially, using systematic mutagenesis analysis of splice sites and branch points in circRNA expression vectors, Starke et al. proved that back-splicing requires canonical splicing signals (*29*). Since back-splicing uses the same splice sites and spliceosome components as canonical splicing, it is reasonable to speculate that back-

splicing and canonical splicing affect each other. Introduction of canonical splice sites in circRNA expression minigenes dramatically decreases the production of circRNAs, indicating the competition between canonical splicing and circRNA biogenesis (*28*). Interestingly, back-splicing and canonical splicing show different responses to the depletion of core components of spliceosomes, where the linear spliced mRNAs get decreased while the circRNA biogenesis gets enhanced (*32*). Increased circRNA production was also observed in rat neurons treated with splicing inhibitor isoginkgetin (*33*).

1.2.2 Transcription and back-splicing

Back-splicing can happen both co-transcriptionally and post-transcriptionally. circRNAs were identified from RNA-seq data of chromatin-bound RNA samples from fly heads and mouse livers, suggesting they are produced co-transcriptionally (*28*). However, in another study trying to identify nascent circRNAs by 4sU pulse labeling, only a few circRNAs can be identified with short labeling periods (10-15 min) (*34*). Increasing the labeling time increases both the number and abundance of nascent circRNAs dramatically, suggesting that circRNAs are produced after the completion of transcription of their cognate pre-mRNAs (*34*). Using circRNA expression vectors, Liang et al. also showed that a functional 3'-end processing signal is required for circRNA formation, suggesting back-splicing may occur post-transcriptionally (*31*). The coupling between transcription and back-splicing has been related to the elongation rate of RNA polymerase II (Pol II). Analysis of nascent transcripts revealed that circRNA-producing genes have a higher transcription elongation rate (TER) than non-circRNA genes (*34*). In *Drosophila*, a low TER of RNA Pol II results in significantly fewer circRNAs (*28*). It is possible that fast transcription enables sequences in downstream introns to pair with upstream introns before upstream introns are spliced out by canonical splicing (section 1.2.4).

1.2.3 Exon-skipping and back-splicing

circRNA is found to be correlated to exon-skipping (*35-40*). In early years, sporadic examples showed that some circRNA-producing genes generate linear transcripts that skip the exons to be circularized (*38-40*). Later, systematic analysis of RNA-seq data in human cells found that the more exons undergo circularization, the less likely these exons are presented in the linear mRNAs from the cognate genes (*35*). Another study showed that circRNAs from the I-band of the titin gene are altered when RNA-binding motif protein 20 (RBM20) is mutated or lost (*41*). Interestingly, the I band region can be skipped, and the skipping is dependent on RRM20 (*41*). However, it remains unclear whether altered circRNA formation in this region results from RBM 20-dependent skipping or not. The correlated skipping transcripts were also identified when using plasmid-based circRNA vectors (*32, 34, 42*). Moreover, in *Schizosaccharomyces pombe*, Barrett et al. showed that circRNA could be produced from an exon-containing lariat intermediate produced by exon-skipping (*36*). However, it should be noted that not all circRNA genes have the corresponding skipped transcripts, neither can all skipped exons produce circRNAs.

When reporting the first example of the correlation between circRNA formation and exon-skipping in the rat cytochrome P450 2C24 gene, Zaphiropoulos proposed two mechanisms that might explain how the two types of transcripts were produced from the same gene (*40*), which are later referred to as the "direct back-splicing" model and the "lariat precursor" model (*36, 37, 43, 44*) (Figure 1.2A and B). In the "direct back-splicing" model, circRNAs are first produced by back-splicing, and the skipped transcripts are formed from the y-shaped intermediate of back-splicing (Figure 1.2A). In the "lariat precursor" model, exon-skipping

happens first, and then circRNAs are produced by back-splicing of the lariat intermediate (Figure 1.2B). The former model is for circRNAs driven by intronic complementary sequences (section 1.2.4), and the latter model is for circRNA genes lacking such sequences (*36, 43*). Despite these, some other mechanisms might also play a role in the correlation between exon-skipping and back-splicing. In *Arabidopsis*, Conn et al. showed that a circRNA from the exon 6 of the *SEPALLATA3* gene promotes the exon-skipping of this exon 6 by forming an R-loop structure (Figure 1.2C) (*45*).

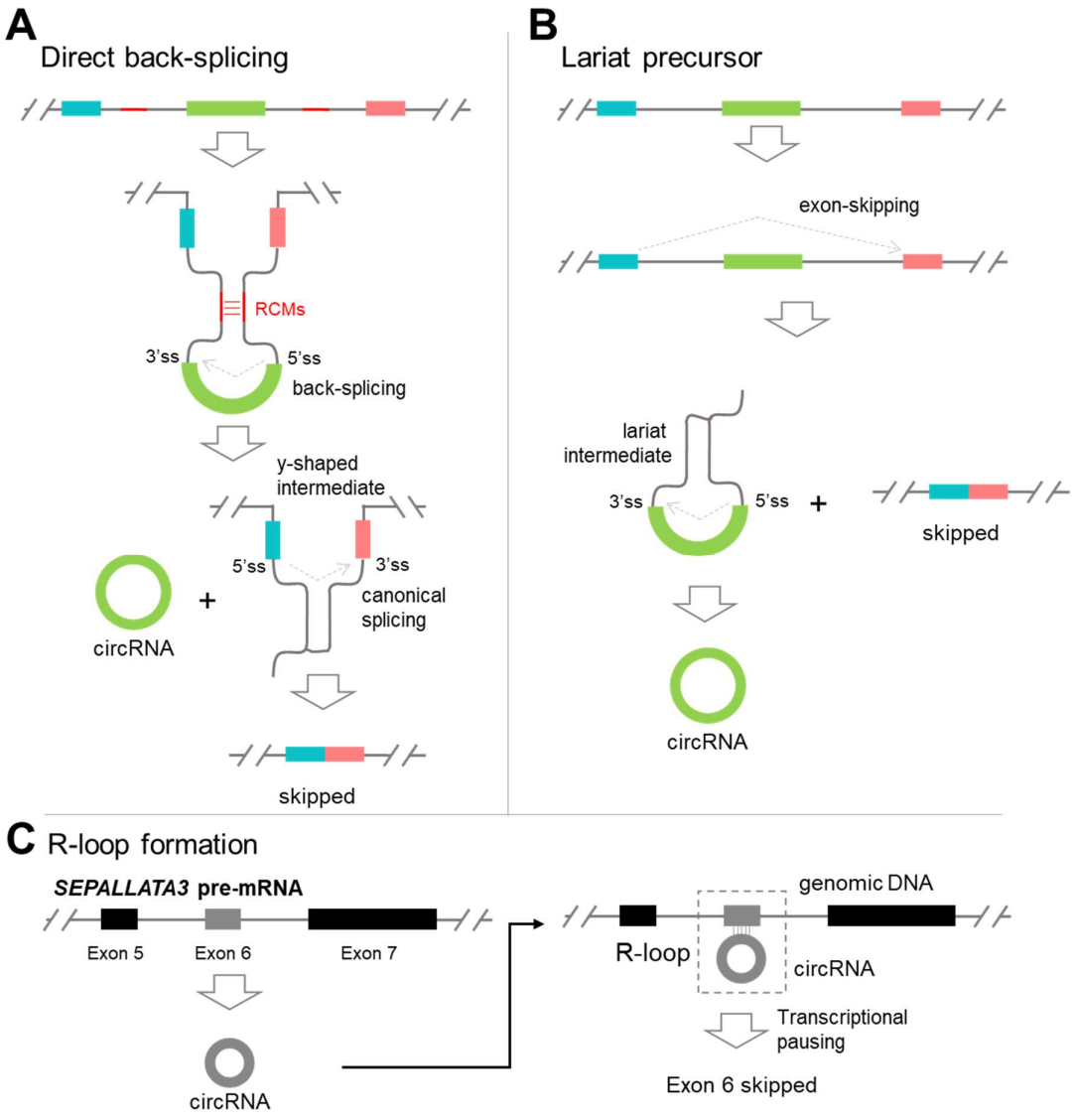


Figure 1.2 Mechanisms involved in the correlated exon-skipping and back-splicing. (A) "Direct back-splicing" model: RCM-driven back-splicing produces a circRNA and a y-shaped intermediate, which is further spliced to form a skipped transcript. (B) "Lariat precursor" model: Exon-skipping happens first to produce a skipped transcript and a lariat intermediate, which serves as a precursor for back-splicing. (C) R-loop formation mechanism: Exon 6 in the *SEPALLATA3* gene produces a circRNA, which interacts with the genomic DNA sequences of this exon in its cognate gene, forming an R-loop structure that promotes the skipping of the exon 6.

1.2.4 *Cis* elements regulating circRNA biogenesis

In the 1990s, when it was first discovered that *Sry* produces mainly circular transcripts, Capel et al. suggested that "inverted repeat" sequences in introns flanking the circle-forming exon could facilitate circRNA production by base-pairing, bringing splice sites into close proximity (*15*) (Figure 1.3A). They proposed this model based on the existence of > 15.5 kb of almost fully complementary sequences flanking the ~ 1.2 kb exon of *Sry* (*46*). In support of this model, subsequent *in vitro* experiments showed that as little as 400 nt of complementary sequences are sufficient to produce *circSry*. The addition of "inverted repeats" to RNA splicing substrates could also promote circRNA formation (*47*). Although very few exons are flanked by such long repeats, this example provided the first *cis* elements that regulate circRNA biogenesis.

Subsequent genome-wide analysis of RNA-seq data in humans revealed that *Alu* repeats, a kind of inverted repeat in the human genome, are enriched in the flanking introns of circRNAs (*37*). Later, Zhang et al. reported that circRNA-forming exons are preferentially flanked by orientation-opposite *Alu* elements, which can promote circRNA formation by base-pairing (*42*). Moreover, non-repetitive but complementary sequences can also enhance circRNA production (*42*). Through deletions of sequences in circRNA-flanking introns of *ZKSCAN1* in an expression vector, Liang et al. showed that as short as 40 nt intronic repeats are sufficient to promote exon circularization (*31*). Similar base-pairing of intronic sequences in circRNA-flanking introns were also identified in other organisms, such as flies (*48*) and nematodes (*49, 50*). More interestingly, introns always contain multiple repeats, which diversifies pattern combinations and results in various circRNAs with different exon contents from the same genes (*42*). This process is called "alternative back-splicing," which, together with other types of alternative splicing, leads to the diversity of circRNA landscape (Figure 1.3) (*22, 51*).

Other *cis* elements from the same loci as circRNA genes can also regulate circRNA biogenesis—for example, the sequences recognized by RNA-binding proteins (RBPs) (section 1.2.5). A notable example is *circMbl* from the *muscleblind* (*mbl*/MBNL1) gene in *Drosophila*, where MBL protein recognizes and binds to the sequences in the *circMbl*-flanking introns and promotes circRNA formation (*28*). In this case, the intron sequences bound by MBL and the MBL protein itself are all *cis* regulating elements for *circMbl*.

Currently, many names have been used to indicate the reverse complementary sequences that promote circRNA formation: inverted repeats (IRs) (*46, 47*), intronic complementary sequences (ICSs) (*2, 34, 44*), and reverse complementary matches (RCMs) (*49, 52*). Since not all such elements are derived from repeated sequences, RCM is used to indicate these sequences in this thesis.

The existence of RCMs has been used to predict circRNA formation with reasonable accuracy (*49, 52*). This model provides a simple, powerful method for circRNA overexpression. Also, the RCM sequences can be potential targets for circRNA knockout/knockdown with minimal effect on the cognate mRNA, for example, knockout of *circGCN1L1* in human cells (*34*), *cia-cGAS* in mice (*53*), *circERBB2* in gallbladder cancer cell lines (*54*), and *circ-E-Cad* in glioma stem cell lines (*55*).

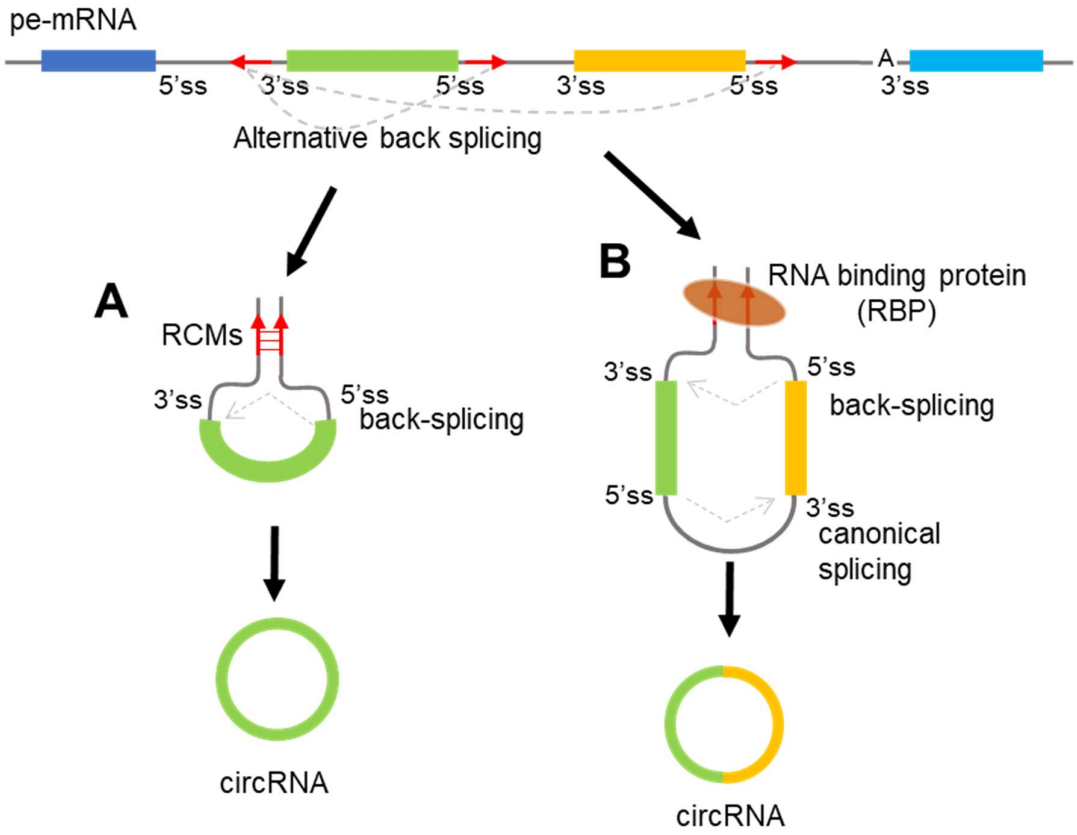


Figure 1.3 circRNA biogenesis is regulated by RCMs and RBPs. (A) Base-paring between RCMs facilitate back-splicing by bringing the splice sites into proximity. Note that different choices between RCM sequences in introns can result in the production of circRNAs with different exon contents, which is called alternative back-splicing. (B) RBPs bind to circRNA flanking introns, regulate back-splicing efficiency.

1.2.5 *Trans* elements regulating circRNA biogenesis

Elements in *trans* can also regulate circRNA formation. Currently identified *trans* elements are all RBPs that regulate back-splicing efficiency (Figure 1.3B). Depending on the properties of each RBP, their roles in circRNA production are different.

As an RNA-editing enzyme, adenosine deaminase acting on RNA-1 (ADAR1) binds to the double-stranded RNA (dsRNA) region and transforms adenosines into inosines (*56*). ADAR1 can interrupt the base-pairing of RCMs and inhibit circRNA formation (*49, 52*). ADAR1's role in circRNA biogenesis is conserved across species, from flies to mice and human cells (*49, 52*). Another dsRNA-binding protein, DHX9, also negatively regulates circRNA biogenesis (*57*). DHX9 has helicase activity; hence it is speculated that DHX9 could unwind RNA pairs formed by RCMs in circRNA-flanking introns and inhibit circRNA formation (*57*). In support of this speculation, loss of DHX9 leads to global upregulation of circRNAs (*57*). A later study also reported that depletion of DXH9 in HEK293T cells promotes the formation of multiple circRNAs from the *SMN* gene (*58*). dsRNA-binding proteins (dsRBPs) can also be beneficial to circRNA formation. In a genome-wide RNAi screening of circRNA regulators, more than one hundred of RBPs were identified, among which the immune factors, NF90/NF110, facilitate circRNA production by direct binding to dsRNA formed by RCMs in nascent pre-mRNAs (*59*).

RBPs with single-stranded RNA (ssRNA)-binding capacity can also regulate circRNA formation by binding to specific RNA motifs. Except being a *cis* element for *circMbl*, MBL protein also promotes *circLuna* formation, serving as a *trans* element in this case (*28*). Splicing factor Quaking (QKI) enhances circRNA production during human epithelial-mesenchymal transition (EMT) (*60*). QKI binds to the ssRNA motif in flanking introns of the exon(s) to be circularized and then dimerizes, which results in enhanced circRNA formation (*60*). Moreover, the insertion of synthetic QKI binding sites into introns is sufficient to induce circRNA formation (*60*). Another splicing factor, Fused in Sarcoma (FUS), can regulate circRNA generation by binding to circRNA-flanking introns in mouse embryonic stem cell (eSC)-derived motor neurons (*61*). In human prostate cancer, the heterogeneous nuclear ribonucleoprotein L (HNRNPL) also regulates circRNA formation by binding to the flanking introns (*62*). Some other RBPs are regulators of circRNAs derived from specific genes. As mentioned before, RBM20 is critical for the formation of a subset of circRNAs originated from the I-band of the titin gene in the mammalian heart (*41*). Sam68, an RBP homologous to QKI, promotes multiple circRNA formation in the *SMN* gene by binding in the proximity of *Alu*-rich regions in *SMN* introns, presumably stabilizing the base pairing of *Alu* repeats (*58*). Very recently, Stagsted et al. reported that splicing factor proline/glutamine rich (SFPQ) is a key regulator of circRNAs with distal *Alu* repeats and long flanking introns by preserving accurate long intron splicing (*63*). RBPs can work together to regulate circRNA formation. Multiple hnRNP (heterogeneous nuclear ribonucleoprotein) and SR (Ser/Arg) proteins function in a combinatorial manner for the efficient back-splicing of *Laccase 2* in *Drosophila melanogaster* (*48*).

The reported circRNA-regulating RBPs are summarized in Table 1.1.

In all, regulation of circRNA production is a complicated and well-tuned process. Given that a substantial number of circRNAs show tissue-/cell type-specific and developmental stage-specific patterns, a tightly regulated mechanism network should exist, which involves *cis* and *trans* elements.

Table 1.1 Summary of RBPs regulating circRNA biogenesis.

RBP	Organism/cell	Targets	Positive/ Negative	Methods	Ref.
Muscleblind	*Drosophila* S2	*circMbl*; *circLuna*	Positive	Overexpression; minigene	(*28*)
ADAR1	Human HEK293; Mouse P19 and human SH-SY5Y	multiple	Negative	RNAi	(*49, 52*)
Quaking	Human mammary epithelial cells	multiple	Mainly positive	RNAi screening; IRES reporter	(*60*)
hnRNPs(Hrb27C, Hrb87F); SR proteins (SF2, SRp54, B52)	*Droshaphila* S2	multiple	Mixed	RNAi	(*48*)
RBM20	Mice	multiple circRNAs from *Tintin*	Positive	Knockout	(*41*)
DHX9	Human FLPinTrex HEK293 cells; HEK293T cells	multiple; circRNAs from *SMN*	Negative	RNAi	(*57, 58*)
FUS	Mice eSC-derived motor neurons	multiple	Mixed	Knockout; RNAi	(*61*)
HNRNPL	Human LNCaP	multiple	Mixed	RNAi	(*62*)
NF90/NF110	Human Hela	multiple	Mixed	RNAi screening, IRES reporter	(*59*)
Sam68	Human HEK293T	circRNAs form *SMN*	Positive	RNAi	(*58*)
SFPQ	Human HepG2 and HEK 293T Cells	Multiple	Mainly positive	RNAi	(*63*)
AUF1	Hepatocellular carcinoma cell lines	*circMALAT1*	Positive	RNAi	(*64*)

1.2.6 circRNA homeostasis

Because of their circular structures without 5'-ends or 3'-ends, circRNAs show high stability and are resistant to degradation by exoribonucleases involved in many RNA decay mechanisms. In line with this, circRNAs show much longer half-lives than their cognate linear mRNAs. Jeck et al. showed that many circRNAs have > 48 hours of half-lives after actinomycin D treatment, much longer than their linear transcripts, which show < 20 hours of half-lives (*37*). By measuring half-lives of circular and linear transcripts from 60 circRNA genes, Enuka et al. reported that circRNAs have a median half-life of 18.7 – 23.7 hours, while the linear RNAs show 4.4 – 7.4 hours of median half-life (*65*). Hence circRNA can accumulate to high levels during development or aging, especially in post-mitotic cells, like neurons (*50, 66-68*). Although the enrichment of circRNAs in brain tissues is always independent of their linear mRNAs (*66, 67*), it is possible that high levels of circRNAs are due to the different decay rates between the linear and circular transcripts.

Most of the current studies focused on the steady-state levels of circRNAs, which is a balance between biogenesis and degradation. As summarized before, circRNA formation is regulated by many factors. However, how circRNAs are degraded is still poorly understood.

Argonaute 2 (AGO2) is an endoribonuclease responsible for the slice of target mRNAs in the RNAi pathway. *CDR1as* can be degraded by AGO2 when miR-671 binds to it (Figure 1.4A and section 1.3.1) (*69*). RNAi is widely used as an effective method to knock down specific circRNAs. The degradation of circRNAs by RNAi/AGO2 requires sequence recognition by miRNAs or short interfering RNAs (siRNAs). Some other endoribonucleases are also involved in circRNA decay without recognizing specific sequences. The endoribonuclease RNase L degrades circRNAs globally when activated upon viral infection (Figure 1.4B) (*70*). Some N^6-methyladenosine (m^6A)-modified circRNAs can be targeted by RNase P/MRP complex, in which m^6A reader protein YTHDF2 is responsible for recognition and HRSP12 makes connections between YTHDF2 and RNase P/MRP complex (Figure 1.4C) (*71*). In a screening of 31 genes with known functions in RNA metabolism, Jia et al. identified that knockdown of GW182 increases circRNA levels in *Drosophila* cells (Figure 1.4D) (*72*). They further showed that the human homolog proteins of GW182 function similarly, suggesting a conserved role of GW182 protein in circRNA decay (*72*). GW182 is a component of the P-body and a scaffold protein of RNAi machinery (*73*). Through domain deletions in GW182, they found that the Ago-binding domain or P-body targeting domains are dispensable, and the Mid domain of GW182 is critical for circRNA degradation (*72*). However, other components, especially the ribonucleases, interacting with GW182 are not known yet. In another study, Fischer et al. reported that ATP-dependent RNA helicase upstream frameshift 1 (UPF1) and its associated endonuclease G3BP1 selectively degrade highly structured RNA molecules, including circRNAs (Figure 1.4E) (*74*).

Since there are no ends in circRNAs, it is likely that endoribonucleases first cut them into linear molecules, and then exonucleases further degrade the linearized molecules. Nevertheless, the circRNA biogenesis mechanisms and decay pathways work together to maintain the homeostasis of circRNAs.

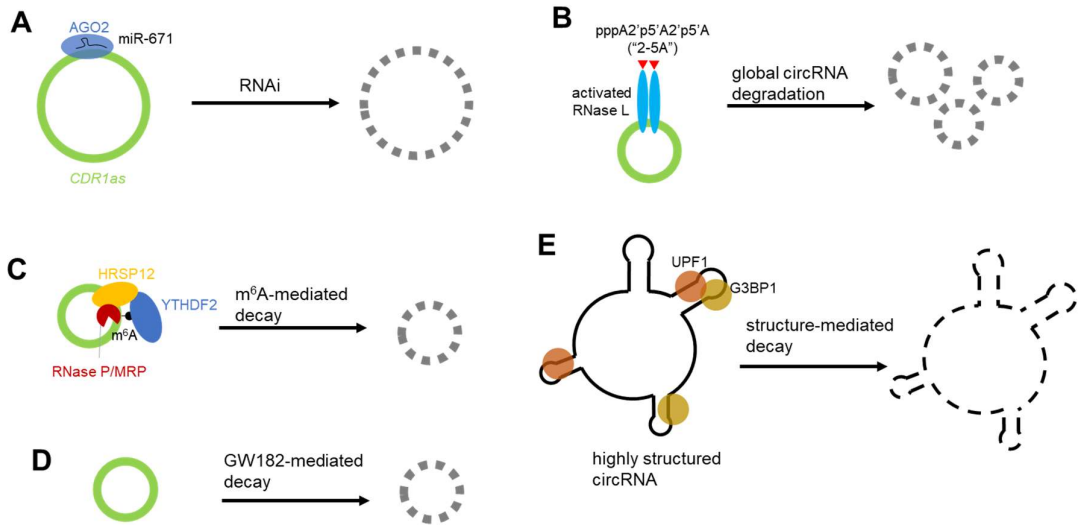


Figure 1.4 circRNA decay pathways. (A) miR-671-mediated degradation of *CDR1as* by AGO2. (B) Global circRNA degradation by activated RNase L. (C) m^{6}A-mediated circRNA decay by RNase P/MRP, in which m^{6}A reader YTHDF2 and adaptor protein HRSP12 are involved. (D) GW182 is involved in circRNA decay. The other components involved are not clear yet. (E) Structured-mediated decay of circRNAs by UPF1 and G3BP1.

1.3 circRNA Functions

Based on the data from CIRCpedia v2 (*75*), hundreds of thousands of circRNAs have been predicted from RNA-seq data from humans, mice, flies, zebrafish, worms, etc. Given that levels of circRNAs are generally low, it is possible that they might be by-products of the splicing process of eukaryotic genes. However, recent findings have suggested that at least some circRNAs play various roles in different physiological environments.

1.3.1 miRNA sponge

It has been proposed that competing endogenous RNAs (ceRNAs) can participate in the regulation of microRNA (miRNA) activity by acting as decoys or sponges through their conserved miRNA binding sites (*76*). Given the high stability of circRNAs, they can serve as miRNA sponges if miRNA-binding sites are in their sequences.

In mouse *circSry* and human *CDR1as*/*ciRS-7*, a considerable number of miRNA binding sites have been revealed. There are 16 binding sites for miR-138 in *circSry* (*77*), while *CDR1as* contains >70 conserved binding sites for miR-7(*77, 78*). What's more, binding sites for miR-7 in *CDR1as* are not all fully complementary, which allows dense binding of miR-7 and at the same time prevents degradation of *CDR1as* through the RNAi pathway (*77*). Consistent with this idea, co-transfection of *CDR1as* with miR-7 significantly reduces the knockdown efficiency of miR-7 target genes (*77*). More interestingly, *CDR1as* also contains a region near-perfectly complementary to miR-671, which enables the circRNA to be cleaved by AGO2 (*69*). Based on these results, Memczak et al. proposed that *CDR1as* could help transport miR-7 to subcellular locations, where miR-671 could trigger the release of miR-7 and the subsequent silencing of miR-7 target genes (*78*).

CDR1as is highly expressed in the excitatory neurons in mice. Piwecka et al. generated a *Cdr1as*-knockout mouse model by deleting the whole ~ 2.9 kb locus coding *Cdr1as*. *CDR1as* knockout mice exhibit dysfunction of excitatory synaptic transmission and abnormal brain function associated with neuropsychiatric disorders (*79*). This phenotype was presumably caused by the loss of *Cdr1as* (*79*). However, although *Cdr1as* represents the most dominant transcript from the deleted locus (*79*), it is also possible that the other transcripts from the deleted genomic sequences are involved in the observed phenotypes, like the linear long non-coding RNA from which *Cdr1as* is produced, or the *Cdr1* gene from the sense strand (*80*). The long non-coding RNA (lncRNA), *Cyrano*, is also involved in the interactions of *CDR1as*, miR-7, and miR-671, forming a well-organized regulatory network in the mammalian brain (*81*).

In addition to *CDR1as* and *circSry*, some other circRNAs were also reported to function through a miRNA sponge mechanism. For example, *circHIPK3*, derived from the second exon of the *HIPK3* gene, regulates cell growth by sponging multiple miRNAs (*82*). *circBIRC6*, which was enriched in undifferentiated human embryonic stem cells (hESCs), suppresses hESC differentiation by attenuating target genes of miR-34a and miR-154 (*83*). circRNAs reportedly showing miRNAs sponge functions are summarized in Table 7.1.

It should be noted that miRNA sponge function may only be limited to a small subset of circRNAs. This is mainly due to the generally low levels of most circRNAs and most annotated circRNAs contain few miRNA-binding sites (*65, 84*). Despite the limited miRNA-sponge capacity of endogenous circRNAs, synthetic circRNAs have been used as effective tools to regulate gene expression by targeting specific miRNAs (*85-87*).

1.3.2 circRNA-protein interaction

circRNAs act as miRNA sponges by sequence recognitions. Similarly, they can also be recognized by RBPs if specific sequences or structures exist, acting as decoys of RBPs or as scaffold structures for RNA-protein complexes.

In the production of *circMbl* in *Drosophila*, multiple MBL proteins bind to flanking introns to promote *circMbl* formation. Also, multiple MBL binding sites exist in *circMbl*, which can sequester MBL protein's other functions (*28*). *circFOXO3* functions as a protein decoy for cell cycle proteins: cyclin-dependent kinase 2 (CDK2) and cyclin-dependent kinase inhibitor 1 (p21), forming a *circFOXO3*-p21-CDK2 ternary complex that arrests the function of CDK2 and p21 (*88*). Another example is *cia-cGAS*, which is highly enriched in the nuclei of human long-term hematopoietic stem cells (LT-HSCs) (*53*). *Cia-cGAS* binds to cyclic GMP-AMP synthase (cGAS), the sensor for exogenous dsDNA for innate immunity, acting as an antagonist to dsDNA, which protects LT-HSCs from interferon-mediated exhaustion (*53*). The binding of circRNA to proteins can be evolutionarily conserved across species. Multiple circRNAs from the conserved BOULE gene bind to heat shock proteins in germ cells of flies, mice, and humans, protecting against stress-induced fertility decline (*89*). There are many other examples of circRNA-protein interactions, which are summarized in Table 1.2.

These identified circRNA-protein interactions are either based on protein co-immunoprecipitation or RNA pulldown, followed by mass spectroscopy for protein identification, which may only reveal a part of the "interactome" of circRNAs.

Table 1.2 Summary of circRNA-protein interactions.

circRNA	host gene	protein	organism/cell type	Ref.
circ-Foxo3	*Foxo3*	CDK2, p21; ID1, E2F1, HIF1α, FAK.	NIH3T3 cells; mouse embryonic fibroblasts	(*88*), (*90*)
circ-Amotl1	Human angiomotin-like 1 gene (*Amotl1*)	PDK1, AKT1	Human and mouse cardiac tissues	(*91*)
circRNA_102171	not mentioned	CTNNBIP1	Thyroid cancer tissues and cell lines	(*92*)
circDNMT1	*DNMT1*	p53, AUF1	Breast cancer cells	(*93*)
circAGO2	*AGO2*	human antigen R (HuR)	Multiple cancer cell lines	(*94*)
circFAT1(e2)	*FAT1*	Y-box binding protein-1 (YBX1)	Gastric cancer cells	(*95*)
circFndc3b	*Fndc3b*	Fused in Sarcoma (FUS)	Mouse cardiac endothelial cell line	(*96*)
circPOK	*Zbtb7a (POKEMON, LRF)*	ILF2/3 complex	Mesenchymal tumor	(*97*)
circACC1	*ACC1*	AMP-activated protein kinase β and γ	HCT116	(*98*)
circNfix	*Nfix*	Ybx1	Cardiomyocyte	(*99*)
cia-cGAS	*D430042O09Rik*	cyclic GMP-AMP synthase (cGAS)	Long-term hematopoietic stem cells	(*53*)
circSKA3	*SKA3*	Integrin β1, Tsk5	Breast cancer	(*100*)
circZNF827	*ZNF827*	hnRNP K,hnRNP L, ZNF827	Cultured neuron cells	(*101*)
circERBB2	*ERBB2*	PA2G4	Gallbladder cancer cells	(*54*)
circFOXK2	*FOXK2*	YBX1, hnRNP K	Pancreatic ductal adenocarcinoma cells	(*102*)
circRHOT1	*RHOT1*	TIP60	Hepatocellular carcinoma cells	(*103*)
circSamd4	*Samd4*	PURA, PURB	C2C12 myoblasts	(*104*)
circPABPC1	*PABPC1*	ITGB1	Human hepatocellular carcinoma	(*105*)
circBoule (multiple circRNAs)	*Boule (Drosophila)*, *Boule* (mice), *BOULE* (Humans)	Heat shock proteins	Germs of *Drosophila*, mice, and humans	(*89*)

1.3.3 circRNAs can be translated

Canonically, translation of mRNA into protein requires the 5' cap (7-methylguanosine, m^7G) structure and 3' poly(A) tail. Since circRNAs lack both structures, it is suggested that eukaryotic ribosomes cannot load onto circRNAs. Ribosome profiling experiments and polysome-circRNA co-precipitation experiments also showed that circRNAs are not likely to recruit ribosomes (*37, 84*). Despite these results, some studies have shown that at least a subset of circRNAs can be translated in a cap-independent manner using internal ribosome entry sites (IRES) (reviewed in (*106, 107*)). RNA circularization forms new sequences in the junctions, which in principle can form new open reading frames or IRES elements that drive the translation.

By RNAi screening, *circZNF609* was found to regulate proliferation in human myoblasts (*108*). Furthermore, *circZNF609* contains an IRES and an open reading frame that shares the same start codon as its linear mRNA. Overexpression of this circRNA in a vector resulted in the production of a protein that localized mainly in the nucleus. Of note, translation of *circZNF609* was of low efficiency (only ~1% as efficient as that of a linear mRNA), with a weak association with polysomes (*108*). In this study, translation from both overexpression vector and endogenous *circZNF609* was verified (*108*). However, a recent study reported that translation of *circZNF209*-overexpression vector is originated from trans-spliced linear transcripts produced by the overexpression vector, which suggests that particular caution is needed when dealing with circRNA translation from expression vectors (*109*).

Ribosome profiling data from fly heads revealed that as many as 122 *Drosophila* circRNAs might be translated (*110*). For example, *circMbl3*, with an IRES element, is able to produce a protein that was verified by mass spectroscopy (*110*).

In addition to IRES, m^6A modification induces protein translation initiation from circRNAs in human cells (*111*). m^6A-driven translation of circRNA is widespread in human cells and is enhanced upon heat shock, indicating a possible role of circRNA-derived proteins in cellular responses to environmental stress (*111*).

Some circRNAs can produce proteins with functions in brain tumors. Dr. Zhang Nu's group reported multiple functional translated circRNAs in glioblastoma (GBM) (*55, 112-115*). Among them, the 254-aa protein (C-E-Cad) is encoded by *circ-E-Cad*, which is derived from exon 7 to exon 10 of the E-cadherin gene *CHD1* (*55*). Interestingly, *circ-E-Cad* is translated in two rounds, which results in a unique 14-aa C-terminal tail in C-E-Cad, formed by a natural frameshift in the second-round translation (*55*). The unique C-terminal tail renders C-E-Cad capable of binding with epidermal growth factor receptor (EGFR) and activating downstream signals (*55*).

It is still not known how frequently a circRNA can be translated and what are the functions of the translated peptides/proteins. The development of bioinformatic tools that predict circRNA translation potential may be helpful to answer these questions (*116-118*). Nevertheless, artificially circularized RNA molecules can bypass RNA sensors and be effectively translated *in vivo* and *in vitro*, serving as expression vectors with potential therapeutic applications (*30, 119, 120*). Recently, a preprint paper reported the use of circularized RNAs for the development of COVID-19 vaccines (*121*).

1.3.4 circRNAs as biomarkers

Intrinsically, circRNAs are resistant to exonuclease and show much longer half-lives than linear RNAs (see section 1.2.6). Moreover, circRNAs have also been identified in human

peripheral blood (*122*) and can be transported to the extracellular fluid in exosomes (*123*). These two characteristics suggest that circRNA can be used as stable molecular biomarkers for disease diagnosis, especially for cancers (reviewed in (*124*)).

1.3.5 Others

Although the majority of circRNAs are derived from exon(s), introns can also be retained in circRNAs, like EIciRNAs and ciRNAs (Figure 1.1 B and D). Unlike the majority of circRNAs that reside in the cytoplasm, EIciRNAs and ciRNAs are mainly located in the nucleus in human cells (*4, 5*). Some EIciRNAs enhance transcription of their parental genes by interacting with U1 small nuclear ribonucleoproteins (snRNPs) and RNA polymerase II in the promoter regions (*4*). Disruption of the interaction between EIciRNAs and U1 snRNPs could reduce the transcription of their parental genes. ciRNAs accumulate in nuclei of human cells and can also regulate transcription of parental genes by acting as positive regulators of RNA polymerase II transcription (*5*).

Some circRNAs show multiple-facet functions. circMALAT1 has an IRES element and 11-nt complementary sequences to the coding sequences of PAX5 (*64*). circMALAT1 binds to PAX5 mature mRNA and at the same time interacts with ribosomes, which results in the inhibition of PAX5 translation, acting as an "mRNA translation brake" (*64*). In the meantime, circMALAT1 functions as the miR-6887-3p sponge to upregulate JAK2 (*64*). The two properties of circMALAT1 work synergistically to promote the self-renewal of hepatocellular cancer stem cells (*64*).

Fusion circRNAs derived from cancer-associated chromosomal translocations can promote cellular transformation, affect cell viability, and confer resistance to treatment in tumor cells (*125*). Dozens of pseudogenes have been identified to be derived from circRNAs in the human and mouse genomes, which could reshape genome architecture by providing additional CTCF-binding sites (*126*).

To summarize, circRNAs play diverse roles in regulating almost every aspect of gene expression, from transcription to splicing and translation. Also, as RNA molecules, circRNAs can interact with DNA sequences, RNAs, and proteins to show different functions. We can expect that more and more functional circRNAs will be identified, especially those with clinical applications.

1.4 circRNAs in *C. elegans*

Currently, only three papers reported the existence of circRNAs in *C. elegans*. Memczak et al. showed that circRNAs in early developmental cell stages (one/two-cell embryo, oocytes, and sperm) of *C. elegans* were specifically expressed (*78*). Ivanov et al. then characterized the profile of circRNAs in major life stages, from which additional ~800 circRNAs were identified (*49*). They found that introns flanking circRNA exons were much longer (~10 fold) than average and that the number of RCMs in flanking introns was also much greater than average (*49*). Cortés-López et al. reported circRNA profiles in *C. elegans* from the fourth larval (L4) stage to day-10 adult stage, in which they found a global accumulation trend of circRNAs during aging (*50*).

1.5 Aim of this thesis

circRNAs are expressed in a tissue-specific manner. Especially, circRNAs are enriched in the brain tissue of almost all the examined organisms, like humans (*52, 127*), mice (*52, 66*), flies (*68*), monkeys (*128*), et al. circRNAs in the brains are more likely to be derived from synaptic genes and are enriched at the synaptic sites (*52, 127*).

As an excellent animal model, *C. elegans* has many features that are advantageous for circRNA research, like a relatively small and well-annotated genome, high efficiency in genome editing, thousands of available mutant strains, many behavior phenotypes for functional studies, etc. In this thesis, I aimed to use *C. elegans* as an *in vivo* model to study the potential functions and regulation of circRNAs, especially in neurons. However, the neuron-specific profile was not available yet in *C. elegans*. Hence the first aim was to provide the first neuronal circRNA profile.

Based on the obtained neuronal circRNA profile, several neuronal circRNAs were knocked out, trying to identify potential functional circRNAs. Unfortunately, no phenotypes were observed. Then I switched my focus to circRNA regulation in *C. elegans* since our understanding of circRNA regulation has been mainly based on *in vitro* systems. Hence, another aim was to understand what factors/elements can regulate circRNA formation in *C. elegans*, providing *in vivo* perspectives to understand mechanisms involved in circRNA formation.

2. Methods

Worm maintenance

C. elegans Bristol N2 strain was used as the wild type. Worms were maintained using standard conditions on Nematode Growth Media (NGM) agar plates with *Escherichia coli* strain OP50 (*129*) at 20°C. For locomotion assay on N2 and *fust-1(csb21)* strain, worms were cultured at 25ºC. New transgenic worms were generated by microinjection with ~ 40 ng/µl plasmid. The strains used in this study are listed in Table 5.9.

Plasmid preparation

fust-1p::fust-1::mRFP: fust-1 genomic fragment containing sequences from 2181 bp upstream ATG to just before stop codon was cloned into the SmaI site of pHK-mRFP vector in frame with mRFP by In-Fusion HD Cloning Kit (Takara). This plasmid was further used to generate the backbone structure containing *fust-1* promoter and mRFP, to which cDNAs of FUST-1 isoforms (isoform a, isoform b, and ΔN) were inserted by In-Fusion (Takara). The mRFP fused FUST-1 cDNA plasmids were used to generate cDNA only plasmids for splicing reporter rescue by removing the mRFP sequences using In-Fusion (Takara). Splicing reporter of *fust-1* exon 5 was prepared by cloning exon 4 to exon 6 into the plasmids provided by Dr. Adam Norris. The full sequences of these plasmids are listed in section7.4, with some annotations highlighted.

Worm synchronization

Worm synchronization was performed by bleaching for large-scale worm preparation (L1 worms for dissociation and RNA extraction). Briefly, worms were washed off plates using M9 buffer when a lot of eggs were laid and most of the worms were gravid adults. The worms were washed with M9 buffer and then bleached in the bleach solution (1 M NaOH, 0.6% (m/v) NaClO) with ~5 minutes of continuous shaking. Then eggs were pelleted and washed three times with 12 ml M9 buffer by centrifuging at 2000 rpm for 0.5 minutes. Finally, the egg pellet was re-suspended in ~5 ml M9 buffer and rocked at room temperature for 17-24 hours to hatch. For small-scale worm preparation (hundreds of worms for chemotaxis, lifespan, aldicarb resistance, and locomotion assay), worms were synchronized by a timed egg-laying approach. Briefly, 10-15 gravid adult worms were placed onto an NGM plate for four hours, and worms were removed after egg-laying. The eggs were then cultured to the desired stage.

RNA extraction

RNA extraction was performed using the Direct-zol RNA MicroPre kit (ZYMO Research) with on-column DNase I (ZYMO Research) digestion according to the manufacturer's protocol. For worm samples, worms were first flash-frozen in Trizol solution (Invitrogen) in liquid N_2 and then homogenized by vortexing with glass beads (φ 0.1 mm) in Beads Cell Disrupter MS-100 (TOMY). For sorted neuron samples, no homogenization is needed. Cell samples in Trizol were directly used for RNA extraction.

L1 worm dissociation

To ~80 µl of L1 worm pellet, 200 µl SDS-DTT solution (200 mM DTT, 0.25% SDS, 20 mM HEPES pH 8.0, 3% sucrose) was added, followed by a 1.5-minute incubation at room temperature. Then, the worm pellet was washed with 5 × 1 ml egg buffer (25 mM HEPES

pH 7.3, 118 mM NaCl, 48 mM KCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$, 0.340 ± 0.005 Osmolarity) and centrifuged at 10,000 ×g for 30 seconds after each wash. The washing steps and centrifugation should be performed quickly so that one round of washing and centrifugation is done in 40-50 seconds. The washed worm pellet was re-suspended in 100 µl pronase (15 mg/ml in egg buffer) from *Streptomyces griseus* (Sigma-Aldrich). Worms were dissociated by periodic mechanical disruption by pipetting for 15 minutes. 200 µl tips were used for mechanical disruption using the method mentioned in Zhang et al.'s protocol ("Pipette the larvae suspension with a 200 µl tip during the digestion. Adjust the pipetting volume to the approximate volume of the suspended pellet. Slowly pull suspended larvae into the pipette tip. Then, press down to force the pipette tip against the bottom of the microcentrifuge tube and slowly eject the contents") (*130*). Do as many times as possible. When most worm bodies were dissociated, 900 µL L-15/FBS medium (10% FBS in Leibovitz's L-15 medium (Gibco), 0.340 ± 0.005 Osmolarity adjusted by sucrose) was added. Cells were collected and washed twice with 1 ml egg buffer by centrifuging at 9600 × g for 5 min at 4ºC. Cells were suspended in the appropriate amount of egg buffer and allowed to sit on ice for at least 30 min. The upper volume of cell suspension was used for FACS. For whole worm control, after dissociation and washing, the cell suspension was put on ice in the whole procedure of sorting.

Fluorescence-Activated Cell Sorting (FACS)

Cell sorting was performed on a FACS AriaII flow cytometer (Becton Dickinson) equipped with a 70 µm nozzle. 2 µm and 3.4 µm polystyrene beads (Spherotech) were used for size calibration. Before sorting, propidium iodide (PI) was added to the cell suspension to a final concentration of 0.2 - 0.5 µg/ml. Then, profiles of dissociated cells from GFP-labeled strains were compared to profiles of cells from N2 worms to exclude auto-fluorescent cells. Sorted cells were collected in 3 ml L-15/FBS medium in a 15 ml conical tube chilled on ice. For RNA extraction, sorted cells and whole worm control samples were collected by centrifugation in a swing-bucket centrifuge at 4400 rpm, 4ºC for 10 min. The supernatant was removed, and 0.3 ml Trizol solution (Invitrogen) was added and stored at -80ºC. For culture, sorted cells were seeded onto a poly-D-lysine coated glass-bottom dish (MatTec) with daily changes of L-15/FBS buffer. Cells were visualized by confocal microscopy (Carl Zeiss, LSM780) with a 60 × oil lens.

Western blot

Protein samples were resolved by SDS-PAGE (5% stacking gel and 12% resolving gel) and transferred to PVDF membrane by the standard protocol (25 V, 30 min) of Trans-Blot Turbo Transfer System (Bio-Rad). After blocking with 5% BSA-PBST (137 mM Sodium Chloride, 10 mM Phosphate, 2.7 mM Potassium Chloride, pH 7.4, 0.1% (v/v) Tween-20, and 5% (w/v) BSA) for 1 hour at room temperature, the membrane was incubated overnight with primary antibody (listed below) at 4ºC. After 3 × 5 min washes in PBST, the membrane was incubated with HRP-conjugated secondary antibody at room temperature for 1 hour. The membrane was washed 3 × 5 min in PBST and then visualized by Amersham ECL Prime Western Blot Detection Reagent (GE Healthcare). Images were taken by Fluorescent Image Analyzer LAS-3000 (FujiFilm) using the chemiluminescence channel. Mouse ANTI-FLAG M2 antibody (F3165, Sigma-Aldrich): 1:2000; alpha-tubulin Mouse mAb (DM1A) (Calbiochem):1:2000; Amersham ECL Mouse IgG, HRP-linked whole Ab (from sheep):1:1000.

Co-immunoprecipitation (Co-IP)

~20,000 L1 FLAG::FUST-1 worms were seeded on a nutrition enriched plate with NA22 *E. coli* (NEP-NA22). After 4-day culture at 20ºC, all bacteria were consumed and most of the progenies were at the L1 stage. Adult worms were removed by filtering through a 30 μm mesh. Three NEP-NA22 plates, which gave ~ 1 million L1 worms, were used for one replicate experiment. Worms were washed with 1 × 10 ml M9 buffer, 2 × 10 ml cold Buffer B70 (50 mM HEPES-KOH (pH 7.4), 70 mM potassium acetate (KAc), 1 mM sodium fluoride (NaF), 20 mM β-glycerophosphate, 5 mM magnesium acetate (MgOAc), 0.1% Triton X-100, 10% glycerol). Worms were then re-suspended in 0.4 ml Buffer B70 supplemented with 2 × cOmplete Proteinase inhibitor cocktail (Roche) and dripped into liquid N_2 with 1 ml pipette tips to form small pearls. Worm pearls were stored at -80ºC. Worm pearls were ground into fine powder in a mortar containing liquid N_2, which was suspended into 1 ml cold Buffer B70 supplemented with 2 × cOmplete Proteinase inhibitor cocktail (Roche) and 5 μl Murine RNase Inhibitor (NEB). Worm lysate was cleared by centrifugation at 20,000 × g for 20 min at 4ºC. 50 μl worm lysate was taken as input samples, in which 40 μl was used for RNA extraction and 10 μl for western blot. 50 μl Dynabeads Protein G (Invitrogen) was coupled with or without 5 μg Anti-FLAG M2 antibody (Sigma-Aldrich), which was then incubated with 400 μl lysate, rotating overnight at 4ºC. On the next day, the lysate-beads slurry was cleared magnetically, and the supernatant was taken for western blot. With tubes on a magnetic tray, the beads were gently washed with 2 × 200 ul Buffer B70. To elute bound RBP complex, 50 μl 50 mM glycine, pH 2.8 was added to the washed beads. After mixing and incubating at RT for 3 min, the supernatant was transferred to another tube containing 5 μl 1 M Tris-HCl, pH 7.5 for pH neutralization. For the 55 μl elution, 44 μl was used for RNA extraction, 11 μl for western blot.

RNA-seq

For RNA-seq of samples from sorted neurons (the sort group) and whole worms (the whole group), libraries were prepared using KAPA RNA HyperPrep kit with RiboErase (HMR) (KAPA biosystems) according to the manufacturer's protocol. The RNA input was 150 ng and fragmentation conditions were 85ºC for 5 min. Barcodes were introduced to each sample using KAPA duel-indexed adapters (KAPA biosystems). The length distribution of each library was determined by TapeStation 4200 (Agilent) using High Sensitivity DNA ScreenTape (Agilent). Libraries were quantified by KAPA library quantification kit (KAPA biosystems) and then multiplexed and sequenced on Illumina Hiseq 4000 platform to obtain 150 nt paired-end reads. For RNA-seq of samples from the L1 stage of N2 and *fust-1(csb21)*, rRNA depletion was performed using Ribo-Zero Plus rRNA Depletion kit (Illumina). Library preparation was conducted using NEBNext® Ultra II Directional RNA Library Prep Kit for Illumina (New England BioLabs) by SQC staff in OIST. Sequencing was performed on NovaSeq 6000 (Illumina) to obtain 150 nt paired-end reads.

Worm sorting

L1 worms of strains with extrachromosomal plasmids were obtained by bleaching and hatching for 17-20 hrs in M9 buffer at room temperature. Fluorescence-positive L1 worms were sorted by Large Particle Biosorter (Union Biometrica). Total RNA was extracted from sorted worms, which was used for RT-qPCR quantification.

Droplet digital PCR (ddPCR)

cDNA was reverse transcribed from 10 ng total RNA using an iScript Advanced cDNA synthesis kit (Bio-Rad). ddPCR was performed by using ddPCR EvaGreen Supermix kit (Bio-Rad) on a QX200 Droplet Reader (Bio-Rad) based on the manufacturer's protocol. Results were analyzed using QuantaSoft software (Bio-Rad).

Real-time PCR

Real-time PCR reactions were performed using soAdvanced Universal SYBR Green Supermix (Bio-Rad) with cDNAs synthesized from iScript Advanced cDNA synthesis kit (Bio-Rad). 20 µl reaction mix with 2 µl cDNA (~1-10 ng) were monitored on StepOnePlus Thermal Cycler (Applied Biosystems) in "fast mode". Cycling conditions: 95ºC, 30', 40 or 45 cycles of 95ºC, 15' and 60ºC, 30'with plate reading, and a final melt curve stage using default conditions. The primers are listed in Table 5.10.

RNase R treatment

Total RNA was treated with or without (Mock) RNase R (2 U/µg) in the presence of Ribolock(2 U/µg) (ThermoFisher Scientific). The reaction was incubated at 37ºC for 30 min. Then RNA was purified with an RNA Clean and Concentrator kit (ZYMO Research) according to the manufacturer's protocol. For fold change quantification, RNA was quantified by Nanodrop and an equal amount of RNA input was used for cDNA synthesis. For northern blot, 20 µg total RNA with or without RNase R treatment was used for loading.

Northern blot

Northern blot was performed using NorthernMax kit (ThermoFisher Scientific), and the probes were labeled by α-^{32}P-deoxycytidine 5'-triphosphate (PerkinElmer) using Random Primer DNA Labeling Kit Ver. 2 (Takara, #6045) according to the manufacturer's protocols. Briefly, RNA samples (10 µg or 20 µg) were resolved in 1% agarose gel by electrophoresis at 5 V/cm in 1 × MOPS buffer for ~2 hours. Then RNA was transferred onto an Amersham Hybond-N+ membrane (GE Healthcare) by capillary blot for 2.5 hours using the transfer buffer supplied in the NorthernMax kit. Transferred RNA was crosslinked by 254 nm UV at 1200 × 100 µJ/cm^2(Analytik Jena CL-1000). Prehybridization was performed in ULTRAHybe buffer at 50ºC for one hour, followed by hybridization with ^{32}P labeled probes overnight at 50ºC. The membrane was washed 2 × 5 min at room temperature using Low Stringency Washing Solution and 2 × 15 min at 50ºC using High Stringency Washing Solution. The membrane was sealed in kitchen wrap and exposed to a phosphorscreen for several hours to overnight, and the signals were detected by Typhoon FLA7000 (GE Healthcare). Quantification of band intensities was performed using ImageQuant software (GE Healthcare). The average intensity of an area with no bands was used as background intensity. The average intensities of each band were subtracted by the background intensity before comparison. Primers used for probe amplification are listed in Table 5.10.

circRNA prediction and RNA-seq data analysis

The DCC pipeline (*131*) was used for the prediction of circRNAs with RNA-seq data. Briefly, raw reads were aligned to reference genome (WBcel235/ce11) using STAR (*132*) (https://github.com/alexdobin/STAR) with the following options: --outSJfilterOverhangMin 15 15 15 15 –alignSJoverhangMin 15 –alignSJDBoverhangMin 15 --outFilterScoreMin 1 --outFilterMatchNmin 1 --outFilterMismatchNmax 2 --chimSegmentMin 15 --chimScoreMin 15 --chimScoreSeparation 10 --chimJunctionOverhangMin 15. Then the output files from STAR, chimeric.out.junction, were used for circRNA annotation with DCC (https://github.com/dieterich-lab/DCC). Predicted circRNAs from DCC were filtered with at least three junction reads in each group. Differential expression analyses of mRNAs and circRNAs were performed using DESeq2 (*133*) package in R with the gene count output from STAR or the BSJ junction count output from DCC, respectively. The plots (PCA plots, boxplots, scatter plots) were generated using the ggplot2 package

(https://ggplot2.tidyverse.org/) and the ggpubr (http://www.sthda.com/english/rpkgs/ggpubr) package in R. The scripts of DESeq2 and plots are in section 7.3.

RCM analysis

RCM analysis in flanking introns of circRNAs or non-circular control exons was performed using IntronPicker and autoBLAST scripts (https://github.com/alexandruioanvoda/) described in (*50*).

Microscopy

Confocal images were obtained using a Zeiss LSM780 confocal microscope, and images were processed using ZEISS ZEN3.1 software. For quantification of GFP and mCherry signals of the *fust-1* exon 5 splicing reporter, confocal images were used without further adjustments. Areas around the neck neurons were selected, and the average intensities of GFP and mCherry (given by ZEN3.1 software) in those selected areas were used for quantification.

Gene ontology analysis

Gene ontology enrichment analysis was performed using WormBase Enrichment Suite web server (*134, 135*) (https://wormbase.org/tools/enrichment/tea/tea.cgi).

Chemotaxis assay

Chemotaxis assays were performed on 10-cm square plates as described previously (*136*).

Lifespan assay

Lifespan assays were performed on NGM-OP50 plates with or without 0.1 mg/ml 5-fluorodeoxyuridine (FUdR). If without FUdR, worms were transferred to new NGM-OP50 plates the other day starting from day 1 adult. Worms were classified as dead and removed if not moving with gentle poking with a worm picker made of platinum wire. Worms bagged, dead with obvious body damage, or burrowed into agar were not considered.

Aldicarb resistance assay

Blank NGM plates containing 1 mM aldicarb (Wako) were used for assays. Young adult worms synchronized by egglaying were used. Worms were transferred to aldicarb plates, and every 30 min, paralysis was checked by gentle poking with a worm picker made of platinum wire. Worms were removed from plates if they were identified as fully paralyzed.

Locomotion assay

Locomotion analysis of day 3 adult worms was performed as described previously (*137*). Briefly, 15 synchronized day 3 adult worms were picked onto a blank NGM plate to get rid of food for ~1 min. The worms were then transferred to another empty NGM plate and locomotion images were recorded for 1 min with five frames per second with the lid on. Images were analyzed using ImageJ and wrMTrck plugin (*138*) (http://www.phage.dk/plugins/wrmtrck.html) to calculate the average speeds. More than 50 worms were recorded. Worms that got lost during recording were not included.

Single worm PCR

Single worm PCR was performed based on an online protocol (https://theolb.readthedocs.io/en/latest/molecular-biology/c-elegans-single-worm-pcr.html).

Mutagenesis by CRISPR-Cas9

Genome editing by CRISPR-Cas9 was achieved based on the protocol from Dokshin et al (*139*) with minor modification. Injection mix (20 µl) containing the following components: Cas9 protein (0.25 µg/µl, Sigma-Aldrich), tracrRNA(0.1 µg/µl, ThermoFisher), crRNA (0.2 µg/µl for one, 0.1 µg/µl each for 2 crRNAs, ordered from ThermoFisher or IDT), injection marker plasmid (PRF4(*rol-6*), 40 ng/µl), KCl (25 mM) and HEPES (10 mM, pH 7.4) was injected to the gonads of young adult worms. After injection, F1 rollers were picked, and their genotypes were checked by single worm PCR to screen the target mutations. For homozygous F1, non-roller F2 progenies were kept as the target strain. For heterozygous F1, homozygous F2s were kept as target strains.

Table 2.1 Injection mix of CRISPR-Cas9 mutagenesis.

Component	Concentration	Solvent	Injection mix	Final concentration
Cas9*	10 µg/µl	50% glycerol	0.5 µl	0.25 µg/µl
tracrRNA*	0.4 µg/µl	10 mM Tris-HCl, pH7.5	5.0 µl	0.1 µg/µl
crRNA1*	0.4 µg/µl	10 mM Tris-HCl, pH7.5	1.4 µl	0.1 µg/µl
crRNA2*	0.4 µg/µl	10 mM Tris-HCl, pH7.5	1.4 µl	0.1 µg/µl
ssDNA	1 µg/µl	10 mM Tris-HCl, pH7.5	2.2 µl	220 ng/µl
Injection marker	300 ng/µl	10 mM Tris-HCl, pH7.5	2.7 µl	~40 ng/µl
KCl	1 M		0.5 µl	25 mM
HEPES	0.2 M, pH7.4		1.0 µl	10 mM
H_2O			5.3 µl	
		Total	20µl	

* Incubate at 37°C for 10 min before adding the other components.

3. Results

3.1 First neuronal circRNA profile from L1 worms

Currently, there are no available neuronal circRNA data in *C. elegans*, mainly due to challenges in obtaining enough neuron samples from the tiny worms that have no obviously compartmentalized "brain" tissue. The most common method to obtain neuron cells from *C. elegans* is by the "labeling-dissociation-sorting" method (Figure 3.1A) (*130, 140-145*), in which target neurons are labeled by fluorescent protein and, after mild dissociation of the worms, labeled neurons are collected by fluorescence-activated cell sorting (FACS). This method can obtain target neurons in high purity and has been used to detect gene expression in single neurons to all the neurons (*140-143*). However, due to the low efficiency of the dissociation (*130, 141*), total RNA obtained from sorted cells is limited. Hence this method is only used for mRNA detection, either by microarray or RNA-seq (*140, 141, 145-147*). By optimization of previous protocols (*130*), I aim to improve the final total RNA yield to hundreds of nanograms for circRNA detection by whole-transcriptome RNA-seq. Worms at the first larval stage (L1) were used to minimize the potential accumulation effect of circRNAs during development.

Of note, results in chapter 3.1 were from the author's published paper with modifications (*148*).

3.1.1 Successful large-scale neuron isolation from L1 worms

Here, using a strain (NW1229) with pan-neuronal green fluorescent protein (GFP) expression, I found that by shortening the time of SDS-DTT treatment (from 2 min to 1.5 min) and washing (from 5 × 1.0 min to 5 × 40 sec) as well as increasing the time of mechanical disruption (from 10 min to 15 min) (Figure 5.1A), cell yield could be improved. After dissociation, the cell suspension was stained with propidium iodide (PI) to label dead/damaged cells and then subjected to FACS. GFP positive singlet cells were sorted (Figure 3.1B). The majority of sorted cells showed GFP fluorescence when observed under a confocal microscope (Figure 3.1C). Some neuron cells kept short neurites after sorting (Figure 5.1B). Consistent with previous findings, neurites can grow out from sorted neuron cells after culturing (Figure 5.1C) (*130, 141*). To further confirm the effectiveness of sorting, the levels of two marker genes (*myo-3* and *unc-64*) were quantified by digital droplet PCR (ddPCR). As expected, the neural syntaxin *unc-64* was highly enriched in the sorted cells, whereas the muscle gene *myo-3* was depleted (Figure 3.1D).

Using this optimized protocol (see Methods), 200 - 500 ng total RNA was obtained from cells sorted from ~ 1.5 - 5 million L1 worms (the sort group). RNA samples from dissociated worms before sorting were also prepared for comparison (the whole group, Figure 3.1A). For RNA-seq, 150 ng total RNA from three independent trials of the sort group and the whole group was used as input for library preparation with ribosomal RNA removal and first-strand cDNA synthesis using random hexamers. More than 45 million 150 nt paired-end reads were obtained for each sample. Differentially expressed genes between the two groups were analyzed by DESeq2 (*133*) (R scripts are in section 7.3.3). Consistent with the ddPCR results (Figure 3.1D), *myo-3* was significantly depleted, while *unc-64* was significantly enriched in the sort group compared with the whole group (Figure 5.1D). The significantly upregulated genes in the sort group were searched in WormExp (*149*)

(https://wormexp.zoologie.uni-kiel.de/wormexp/) to identify whether these genes overlap with previous results of neuronal genes. As expected, the resulted top three datasets were all pan-neural enriched genes determined by microarray analysis of sorted neurons (Figure 5.1E) (*146, 147*), indicating the RNA-seq results from the sorted samples successfully revealed the gene expression pattern in the neurons.

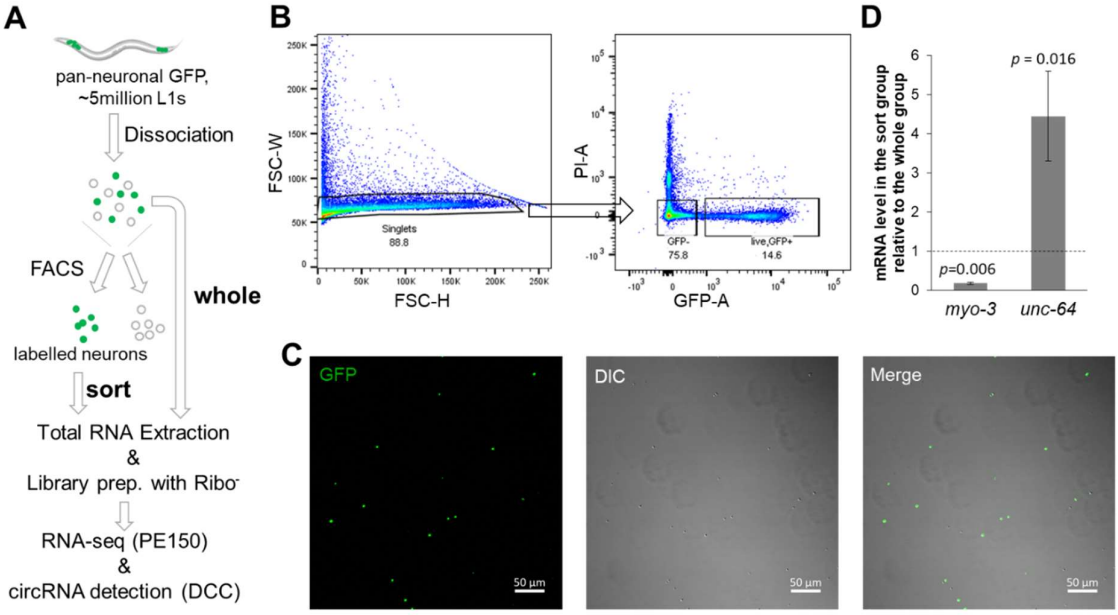


Figure 3.1 Large-scale neuron isolation from *C. elegans* for circRNA detection. Modified from (*148*). (A) Workflow of neuron isolation and circRNA detection by RNA-seq. (B) Gating strategy for FACS: Forward scatter width (FSC-W) was plotted against forward scatter height (FSC-H) to select singlet cells (88.8%), which were then used for the selection of GFP-positive and PI-negative cells (14.6%) for sorting. (C) Confocal images of sorted GFP-positive neurons. Scale bar: 50 µm. (D) ddPCR results showing the relative levels of two genes (*myo-3* and *unc-64*) in the sort group compared with those in the whole group. Error bars stand for standard deviations of three biological replicates. *P* values are ratio paired *t*-test.

3.1.2 circRNA annotation is of high accuracy

circRNA annotation from the RNA-seq data was carried out using the DCC pipeline (*131*). Prior to filtering, 6475 circRNAs were identified with at least one back-spliced junction (BSJ) read across six samples, with 4786 novel circRNAs when compared with circRNAs of *C. elegans* in two databases: circBase (*150*) and CIRCpedia v2 (*75*) (Figure 5.2A). The results were filtered with a cutoff of at least three BSJ reads in each group, which gave 1452 circRNAs derived from 1010 genes and 29 not-annotated loci (Figure 3.2A). Of the filtered circRNAs, the majority of the identified BSJ reads were from exon-to-exon joining (Figure 5.2B). The filtered circRNAs were compared with a published dataset of circRNAs in aging worms (*50*), which showed 450 overlapped circRNA (Figure 5.2C). The novel circRNAs identified in my dataset were mainly from the sorted group, suggesting that sequencing from sorted neuron samples was helpful to identify circRNAs that may not be detected using whole-worm samples.

Two strategies were used to validate the annotated circRNAs by DCC: 1) Amplification of BSJ sequences by RT-PCR using divergent primers followed by Sanger-sequencing (Figure 3.2B). Eighteen out of 19 selected circRNAs, including seven novel circRNAs, were confirmed with the BSJ sequences (Figure 5.2D). 2) Enrichment quantification by RT-qPCR after RNase R treatment. Since there are no ends in circRNAs, they often show resistance to degradation after treatment with RNase R. As expected, while two linear mRNAs (*pmp-3* and *cdc-42*) were depleted after RNase R treatment, all the circRNAs were enriched (Figure 3.2C). The resistance to RNase R was also confirmed by northern blot, which showed that while linear transcript was degraded after RNase R treatment, circRNA from *Y20F4.4* was still detected (Figure 3.2D). Together, these results showed that circRNA annotation is of high accuracy.

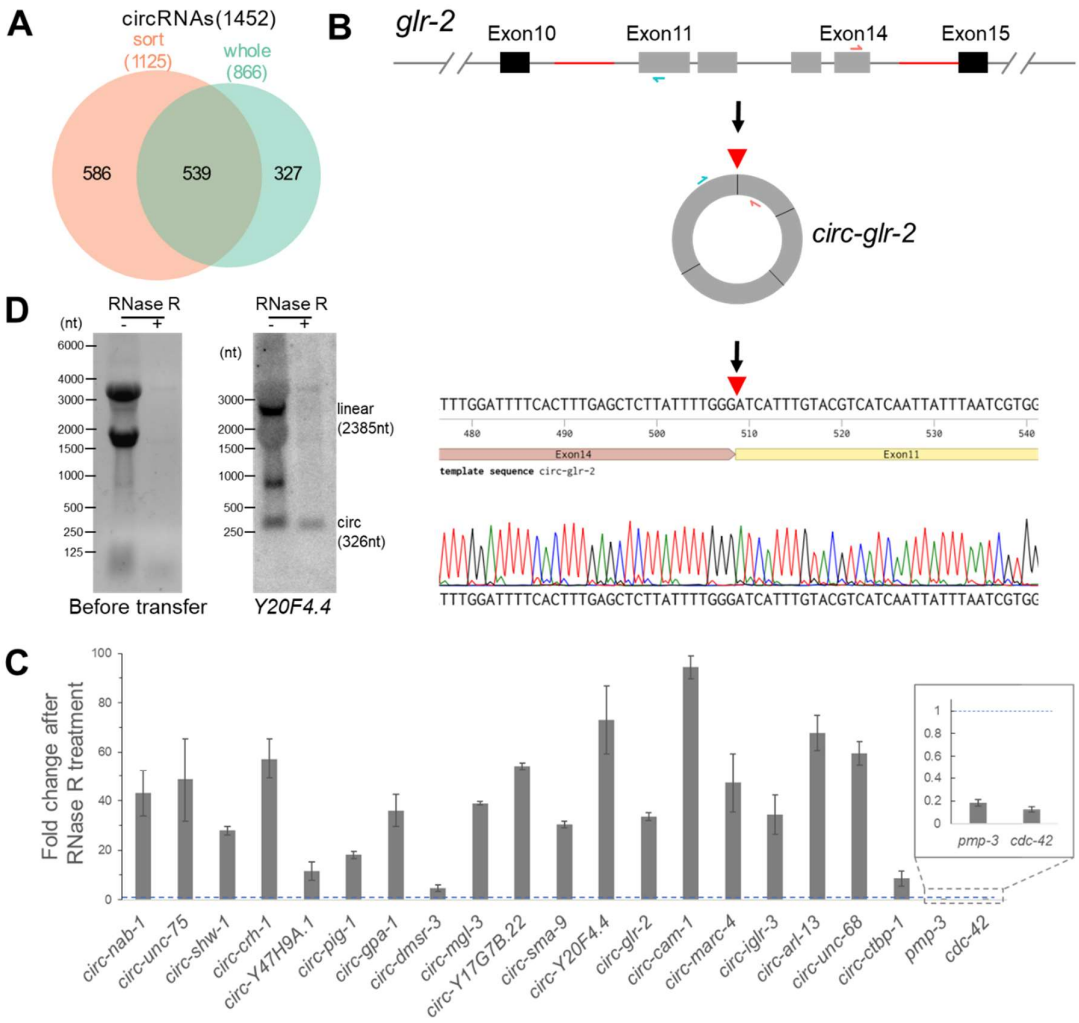


Figure 3.2 circRNA annotation and validation. Modified from (*148*). (A) Overlap of circRNAs detected in the "sort group" and the "whole group." (B) Scheme showing amplification of back-splicing junction of a circRNA from *glr-2* using divergent primers. Amplified sequences are confirmed by Sanger sequencing. The red triangle denotes the joint site. (C) RT-qPCR results of the fold changes of circRNAs and two linear mRNAs (*pmp-3* and *cdc-42*, inset) after RNase R treatment. RNA samples from L1 worms were used. The blue dashed line shows one-fold change. Error bars are the standard deviations of three biological replicates. (D) Northern blot detection of *Y20F4.4* transcripts in total RNA (20 µg) without or with RNAse R treatment, using probes that hybridize to both linear and circular transcripts. RNA samples from L1 worms were used. Left: total RNA with or without RNase R treatment resolved on 1% Agarose gel, stained by EtBr. Right: detection of *Y20F4.4* transcripts after membrane transfer.

3.1.3 circRNAs are highly expressed in the neurons of *C. elegans*

Of the 1452 circRNAs, more circRNAs (1125/1452) were found in the sort group, with 539 identified in both groups and 586 only in the sort group (Figure 3.2A). Gene ontology (GO) enrichment analysis of the circRNA-producing genes showed that terms related to the neuronal functions were significantly enriched, including neurogenesis, synaptic signaling, etc. (Figure 3.3A). Next, the abundances of circRNAs in the sort group and the whole group were compared to check whether circRNAs were highly expressed in the neurons of *C. elegans* or not. TPM (transcripts per million reads) values were used for comparison. The principal component analysis (PCA) plots of both linear mRNAs and circRNAs showed a clear separation between the two groups (Figure 5.3A and 5.3B), suggesting different expression profiles between them. For all the circRNAs in both groups, circRNAs in the sort group showed significantly higher TPM values than in the whole group (Figure 3.3B, $p <$ 2.2e-16, paired Wilcoxon test), indicating circRNAs were enriched in the sort group. The same trend was also observed for the shared 539 circRNAs in both groups (Figure 3.3C, $p =$ 4.7e-14, paired Wilcoxon test).

Next, differentially expressed circRNAs between the sort and the whole group were analyzed, trying to identify neuron-enriched circRNAs. Using BSJ read numbers as input for DESeq2, differential expression analysis of circRNAs between the sort group and the whole group was performed (R scripts in section 7.3.4). With adjusted p value < 0.05, 31 circRNAs were found significantly enriched, and 35 circRNAs were significantly depleted in the sort group (Figure 3.3D). I then asked whether these circRNAs were also derived from neuronal genes or not. The fold changes of circRNAs between the sort and the whole group were plotted against the fold changes of their cognate linear mRNAs. Here, a cutoff of baseMean (given by DESeq2) bigger than 3 was used, which contained 268 circRNAs, including all the significantly differentially expressed circRNAs (Figure 3.3E). The results showed a strong positive correlation (Figure 3.3E, Pearson's correlation coefficient $R = 0.72$, $p <$ 2.2e-16), which indicated that at the L1 stage of *C. elegans*, neuronal circRNAs were more likely to be derived from neuronal genes. When all circRNAs were considered, a moderately strong positive correlation was still observed (Figure 5.3C, Pearson's correlation coefficient $R =$ 0.51, $p <$ 2.2e-16).

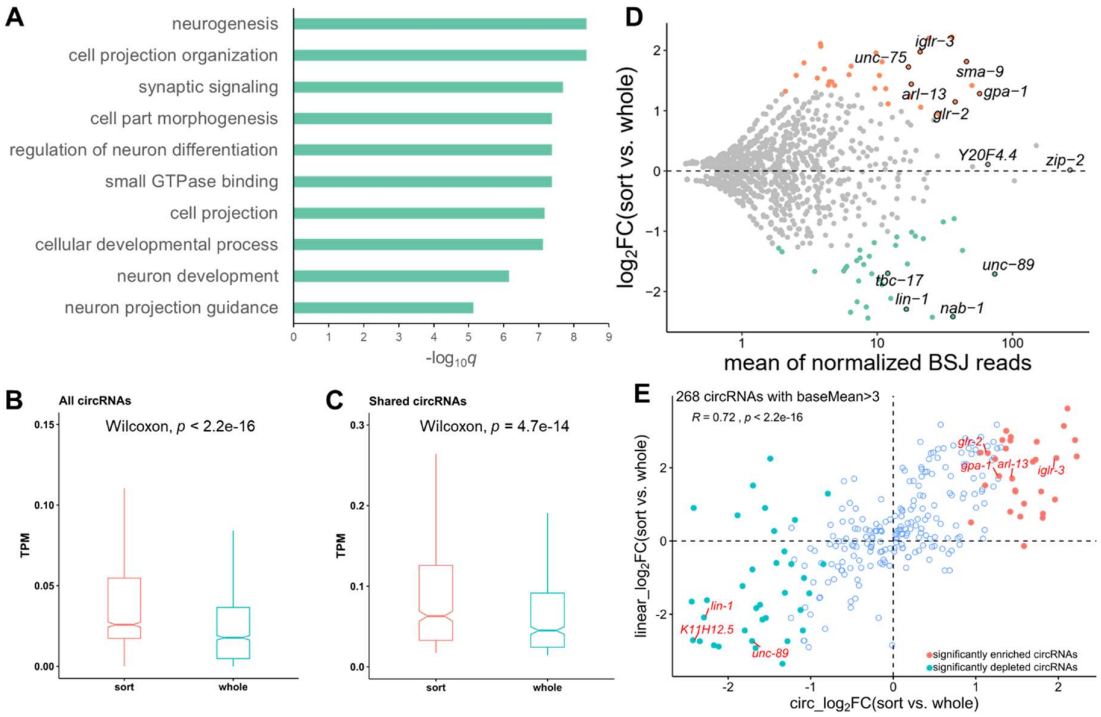


Figure 3.3 circRNA profile in the neurons of *C. elegans*. Modified from (*148*). (A) Top 10 enriched gene ontology (GO) terms of circRNA-producing genes. (B, C) TPM (transcripts per million reads) comparison of all circRNAs (B) and shared circRNAs (C) between the “sort group” and the “whole group”; *p* values are paired Wilcoxon test. (D) Differentially expressed (DE) circRNAs between the sort group and the whole group. Significantly DE circRNAs are highlighted by colors. Gene names of some circRNA genes are labeled. (E) Scatter plot showing the fold changes of 268 circRNAs with baseMean > 3 versus fold changes of their corresponding linear mRNAs. The Pearson correlation coefficient (*R*) and *p* value (*p*) are shown. Significantly differentially expressed circRNAs are shown by colored dots. Names of several circRNA genes are labeled.

3.2 RCMs promote both back-splicing and exon-skipping, simultaneously and directly

Of note, except for results in Chapter 3.2.3, all results in Chapter 3.2 were from the author's published paper with modifications (*148*).

3.2.1 RCMs are abundant in circRNA-flanking introns

Next, features of circRNA-flanking introns were analyzed. Basic local alignment search tool (BLAST) was used to identify RCMs between each pair of flanking introns using the autoBLAST scripts (*50*). Similar to previous findings (*49, 50*), introns that flank circRNA-producing exons were much longer than average, and much more RCMs were identified when compared with flanking introns of non-circular exon controls (exon 2 and exon 8) (Figure 3.4A and B). I also defined best-matched RCM, which is the top one hit with the highest "bit score" in the BLAST results of each pair of introns. Lengths of the best-matched RCMs in circRNA introns were also much longer than those in introns flanking control exons (Figure 3.4C).

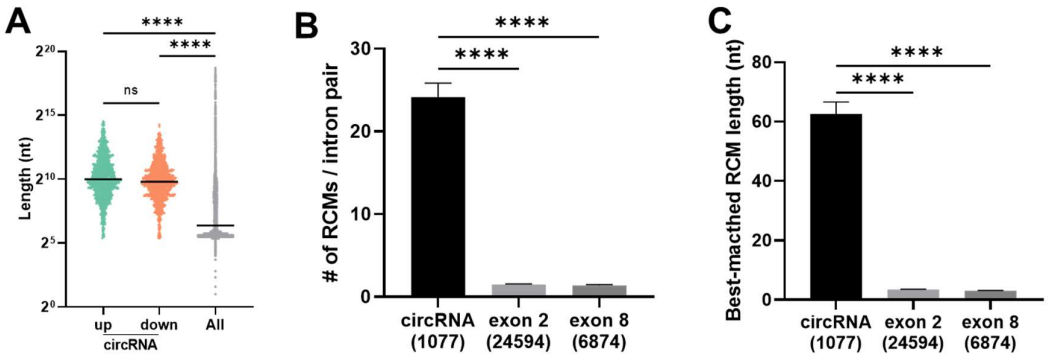


Figure 3.4 RCM analysis of circRNA-flanking introns. Modified from (*148*). (A) Length distributions of introns flanking circRNA-producing exon(s), compared with the lengths of all introns. Average lengths were shown. (B, C) The average number of RCMs (B) and the average length of best-matched RCMs (C) in one pair of circRNA-flanking introns compared with those in non-circRNA control exons (exon 2 and exon 8). Values are shown as mean ± SEM. Numbers in the brackets are numbers of intron pairs used for analysis. *p* values are from the Kruskal-Wallis test with Dunn's post-hoc test for multiple comparisons. ****, $p < 0.0001$.

3.2.2 RCM-deletion is a good way to generate circRNA knockout mutant

Although a previous study showed that RCMs could predict the existence of circRNAs (*49*), the role of RCMs in circRNA formation has not been experimentally confirmed in *C. elegans*. Here, six circRNA genes with RCMs in flanking introns were selected, and one RCM in each gene was deleted using CRISPR-Cas9 (Figure 3.5A). Which RCM is chosen for deletion depends on its position in that intron and the existence of highly specific gRNA sites around RCM sequences. Two guide RNAs (gRNAs) that bracket the target RCM were used for each deletion. A 60-nt single-stranded oligo DNA (ssODN) was used as the repair template, with 30 nt homologous sequences in each end. The gRNA sequences and recombinant single-strand oligo DNAs used for RCM deletions are listed in Table 5.11. For example, in *glr-2*, the downstream RCM extends to the 3' splice site, so only the RCM in the upstream can be used for deletion (Figure 3.5B). RNA-protein complex containing Cas9 protein, tracrRNA, two guide RNAs (gRNA1 and gRNA2) was injected into the gonad of wild-type worms, together with a 60 nt ssODN and an injection marker plasmid (Table 3.1). The injection marker-positive F1 worms were picked on separate plates, whose genotypes were checked by single worm PCR after laying eggs. Six out of the ten picked F1 worms showed different genotypes than wild-type N2 worms (Figure 3.5C), of which four showed the expected band sizes, either homozygous or heterozygous. In order to get heterozygous progeny with designed deletion, single F2 worms were picked from F1-6, which showed a heterozygous pattern, and their genotypes were checked again after egg-laying. As expected, two out of 10 picked F2 worms showed homozygous genotypes (Figure 3.5D). Their sequences were confirmed as the desired deletions precisely at the designed sites (section 7.1 Sequence alignments). Using the same strategy, I deleted one of the RCMs in all the six circRNA genes (Figure 5.4A). The genotype screenings of F1 and F2 worms in the genome deletions of the other five circRNA genes were shown in Figure 5.5. The coordinates and lengths of deleted sequences are shown in Table 3.1.

After obtaining these RCM-deleted strains, circRNA expression was checked. As expected, all the circRNAs were either undetectable or reduced to a very low level after the removal of one RCM in the flanking introns (Figure 3.5E and Figure 5.4B), proving that RCMs in *C. elegans* vigorously promote, if not required for, circRNA formation. Of note, in some of the chosen genes, the linear mRNA levels were altered in RCM-deleted strains.

Table 3.1 Positions and lengths of deleted sequences in circRNA genes. Adopted from (*148*).

Gene name	Coordinates	Upstream/Downstream	Deleted length (bp)
glr-2	chrIII: 7142139 - 7142523	Upstream	385
gpa-1	chrV: 11176808 - 11177313	Upstream	506
unc-75	chrI: 11592753 - 11593798	Upstream	1046
arl-13	chrI: 2066176 - 2066626	Upstream	451
iglr-3	chrI: 2088411 - 2089756	Downstream	1276
Y20F4.4	chrI: 2034765 - 2035642	Downstream	878

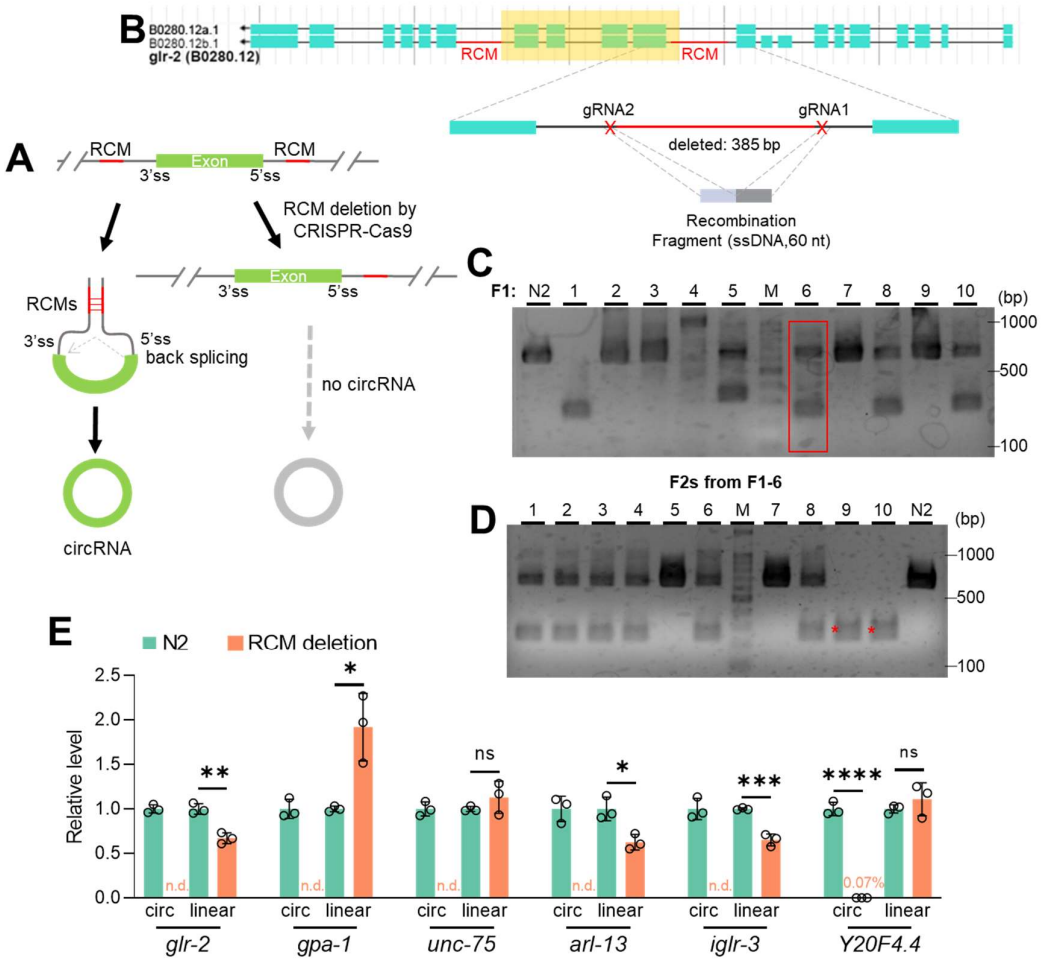


Figure 3.5 RCM deletion by CRISPR-Cas9 to disrupt circRNA formation. Modified from (*148*). (A) Schematic plot showing that RCMs promote circRNA production and the strategy to delete one of the RCMs by CRISPR-Cas9. (B) Upstream RCM deletion in *glr-2*. Positions of RCMs and gRNAs are shown. (C) Genotype screening of F1 progenies from injected P0 worms by single worm PCR. The red rectangle indicates the strain used for F2 screening. (D) Genotype screening of F2 progenies from F1-6. *: Kept as target strains. (E) Quantification of linear mRNA and circRNA in wild-type N2 strain and RCM deletion mutant strains of six circRNA genes. Error bars are the standard deviations of three biological replicates. n.d.: not detected (Ct values not determined or bigger than those in no-template controls). Two-tail student's *t* test. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$. ns: not significant

3.2.3 circRNA knockout strains did not show any apparent phenotypes.

After obtaining these circRNA knockout (KO) strains, a phenotype search was conducted, trying to identify potential functional circRNAs. Another *circ-zip-2* KO strain, *zip-2(ix270)*, was made by deleting the whole intron (intron 1, 130 nt) upstream of the circRNA-producing exons (Table 3.2). Two double circRNA KO strains were also prepared: *gpa-1(ix265); zip-2(ix270)* and *Y20F4.4(ix269); zip-2(ix270)*. The circRNA KO strains used for phenotype search are listed in Table 3.2. All these circRNAs are either significantly enriched or highly expressed in the neurons (Figure 3.3D). Hence the assays used for phenotype search were related to neuronal functions, like chemotaxis assay, locomotion assay, lifespan assay, and aldicarb resistance assay.

Table 3.2 circRNA knockout strains used for phenotype search.

Allele name	Note
arl-13(ix262)	RCM deletion as designed site
glr-2(ix264)	RCM deletion as designed site
gpa-1(ix265)	RCM deletion as designed site
unc-75(ix266)	RCM deletion with random insertions, progeny from F1-4
unc-75(ix267)	RCM deletion as designed site, progeny from F1-6
iglr-3(ix268)	RCM deletion as designed site
Y20F4.4(ix269)	RCM deletion as designed site
zip-2(ix270)	Deletion of whole upstream intron (intron 1)
gpa-1(ix265); zip-2(ix270)	Double mutation
Y20F4.4(ix269); zip-2(ix270)	Double mutation

I started with chemotaxis assays based on our lab's previous experience (*136*). 10-cm square plates were used, where two spots of odor and control were put on the left side and right side symmetrically (Figure 3.6A). Synchronized worms were placed in the middle and left for 10-min crawling, after which worms were killed by chloroform, and chemotaxis index (CI) was calculated by the numbers of worms on the order side and the control side (Figure 3.6A). 1-propanol (20%) was first used as an attractive odor. However, none of the tested strains showed any significant differences compared with the wild-type N2 strain (Figure 3.6B, C, and D). 5% 1-propanol and 0.01% diacetyl were also tested, which did not show any difference, either (Figure 5.6A, B, and C).

Next, I tried locomotion assays. Since circRNAs accumulate during aging (*50*), knocking them out may affect the locomotion ability when worms get old. The moving speeds of worms at different ages were measured using a previously reported protocol (*137*). In the first trial, two *unc-75* strains, CB950 and RM2005 (obtained from CGC), with disturbed protein functions were used as positive controls for locomotion speeds. While CB950 and RM2005 showed significantly lower moving speed, both average speed and maximum speed, from day 1 adult stage to day 4 and day 7 adult stage, only *unc-75(ix266)* showed a slightly significant difference in average speed at day 7 adult stage (Figure 3.7A and B). However, in another independent trial with some other circRNA KO strains, *unc-75(ix266)* did not show any difference in either average speed or maximum speed (Figure 5.7A and B). Locomotion speeds of double circRNA KO strains were also tested together with single circRNA KO strains and wild-type N2 strain. None of them showed any significant difference in moving speed at day 7 adult stage (Figure 5.7C and D).

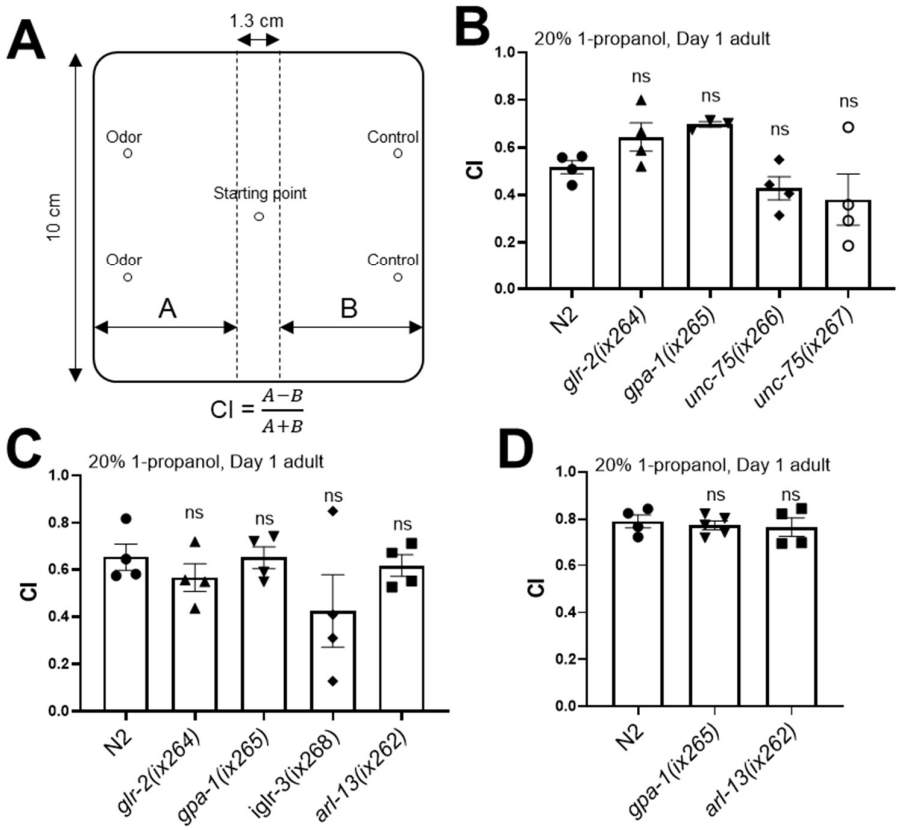


Figure 3.6 Chemotaxis index (CI) towards 20% 1-propanol comparisons between wild-type N2 strain and circRNA KO strains. (A) Scheme showing the 10-cm plate used for chemotaxis assay. Positions of odor and control, the starting point of worms are labeled. (B, C, and D) Independent results performed on different days with different strains. Day 1 adult worms were used. Error bars are standard deviations. ns, not significant, one-way ANOVA with Dunnett's multiple comparisons with N2 group.

Further, aldicarb resistance assay and lifespan assay were also tested (Figure 5.8 and Figure 5.9). In the aldicarb resistance assay, differences were observed in two independent trials between *zip-2(ix270)* and N2 (Figure 5.8A and B, Table 5.1 and 5.2). However, in another separate trial blind to the genotypes, no difference was observed (Figure 5.8C, Table 5.3). For lifespan results, no stable differences in lifespan were observed in the tested circRNA KO strains (Figure 5.9, Table 5.4, 5.5, and 5.6).

In summary, none of these circRNA KO strains showed any obvious or stable phenotypes in the tested assays, suggesting these circRNAs were not involved in these processes.

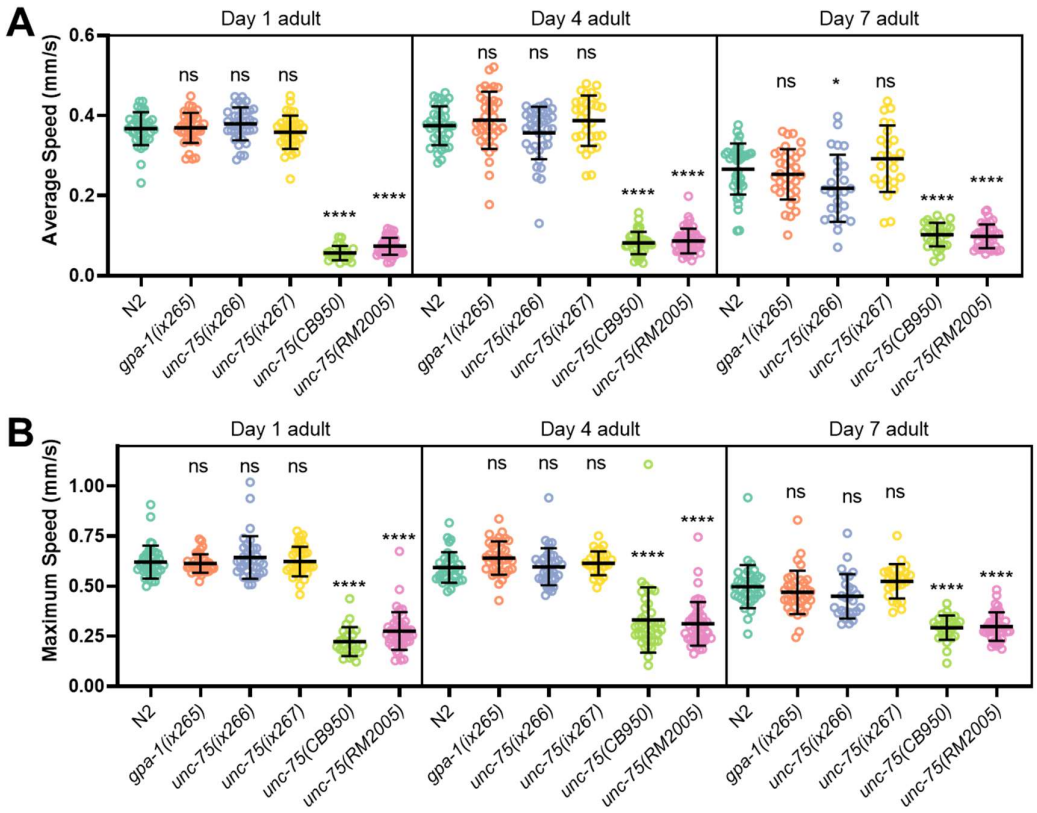


Figure 3.7 Locomotion speed comparisons in different strains at different ages. (A) Average speed comparisons. (B) Maximum speed comparisons. Error bars are standard deviations. *, $p < 0.05$; ****, $p < 0.0001$; ns, not significant, one-way ANOVA with Dunnett's multiple comparisons with N2 group.

3.2.4 Transcripts that skip circRNA-forming exon(s) were identified in some circRNA genes

circRNA production has been correlated with exon-skipping that skips the circularized exon(s) (*35, 38-40*) (Figure 3.8A). In the RNA-seq data of sorted neurons and whole worms, reads mapped to the skipped junctions can be identified in some circRNA genes (Figure 3.8B and Figure 5.10A), suggesting the existence of skipped transcripts. For *zip-2* and *Y20F4.4*, RT-PCR using primers that bracket circRNA-producing exon(s) gave two bands, of which the longer ones were full-length transcripts, and the shorter ones were confirmed to be the skipped transcripts (Figure 3.8C and Figure 5.10C). For some other circRNA genes, the skipped transcripts could be amplified in two-round PCR cycles, in which the corresponding skipped transcripts were gel-cut purified after first-round PCR, which were used as the templates for a second-round PCR (Figure 5.10B and C). In total, skipped transcripts were confirmed in six out of seven chosen circRNA genes, in which only *iglr-3* gave poor sequencing results of the amplified band (Figure 5.10C).

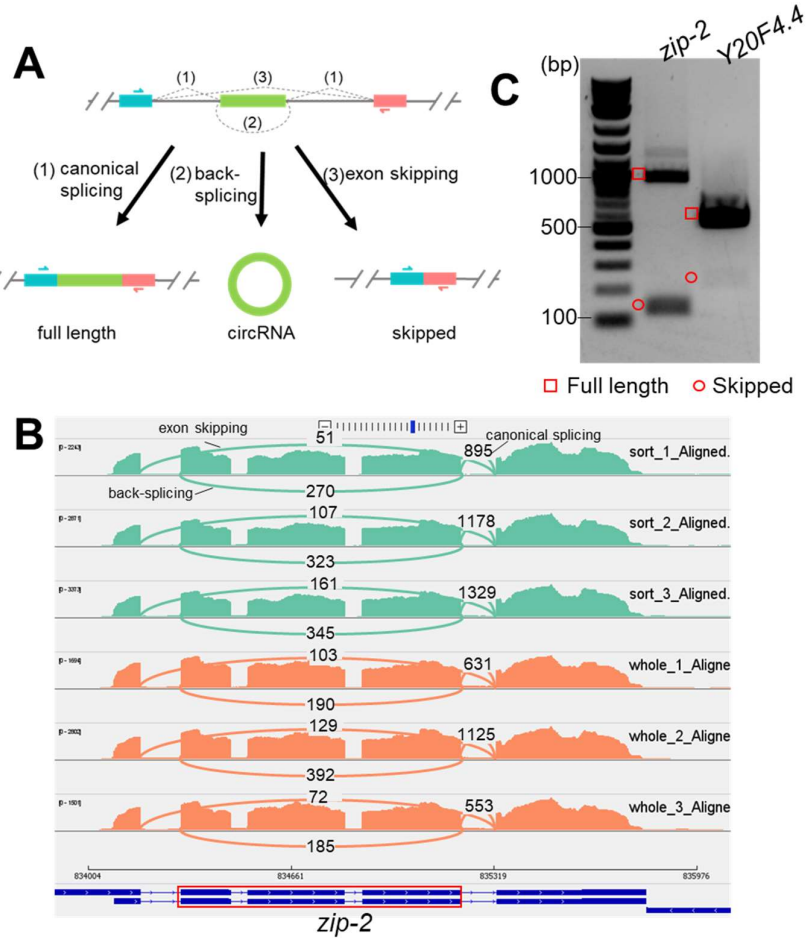


Figure 3.8 Identification of skipped transcripts in circRNA genes. Modified from (*148*). (A) Illustration of three types of transcripts in a circRNA gene. Positions of primers used for amplification of both the full-length and the skipped transcript are shown. (B) Sashimi plot showing numbers of reads aligned to the skipped junction (exon-skipping), the back-spliced junction (back-splicing), and the canonical splicing junction between exon 4 and exon 5 (canonical splicing) in *zip-2*. Exons in the red rectangle are to form *circ-zip-2*. (C) Detection of both full-length and the skipped transcripts by RT-PCR in *zip-2* and *Y20F4.4*. RNA samples from the L1 stage of N2 worms were used. Amplified bands were resolved on a 2% agarose gel.

3.2.5 RCMs simultaneously promote back splicing and exon skipping

Previous studies have shown that complementary sequences in different introns regulate mutually exclusive splicing (*151-153*). Since the correlated exon-skipping and circRNA formation use the same pair of introns, it is possible that RCMs in these introns also affect exon-skipping.

In *Y20F4.4*, deletion of the downstream RCM abolished circRNA formation and did not affect the splicing of the full-length transcript; however, the skipped transcript seemed to be affected (Figure 3.9A). Quantification of the three transcripts in *Y20F4.4* showed that both the circRNA and the skipped transcript levels were strongly decreased after removing the downstream RCM sequences (Figure 3.9B). In *arl-13*, deletion of the upstream RCM eliminated circRNA formation, but the skipped transcript could still be detected (Figure 3.9C). Quantification results showed that the skipped transcript was downregulated in the RCM-deleted strain (Figure 3.9D).

In *zip-2*, two pairs of perfectly matched RCMs, 7 nt and 13 nt in length, respectively, were identified (Figure 3.9E and Figure 5.11A). The RCM sequences in intron 1 and intron 4 were referred to as RCM1 and RCM2, respectively. Deletions of the RCMs were achieved by CRISPR-Cas9, which gave two strains: *zip-2(ix310)* and *zip-2(ix311)* (Table 5.7, Figure 5.11 and Figure 5.12B and C). Canonical splicing of intron 1 and intron 4 was not affected by RCM deletions (Figure 3.9F, Exon 1-2 & Exon 4-5). However, although the circRNA and skipped transcript can be detected in the RCM-deleted strains, their production seemed not as efficient as in the wild-type N2 strain (Figure 3.9F, circ & skip). Levels of the three transcripts of *zip-2* (circular, skipped, full-length linear) were quantified by RT-qPCR. The results showed that while full-length linear *zip-2* was only slightly affected, both the circRNA and the skipped transcript were dramatically reduced in both RCM-deleted mutant strains (Figure 3.9G and H).

Together, these findings suggest that RCMs in the flanking introns of circRNA-producing exon(s) also promote the skipping of these exon(s).

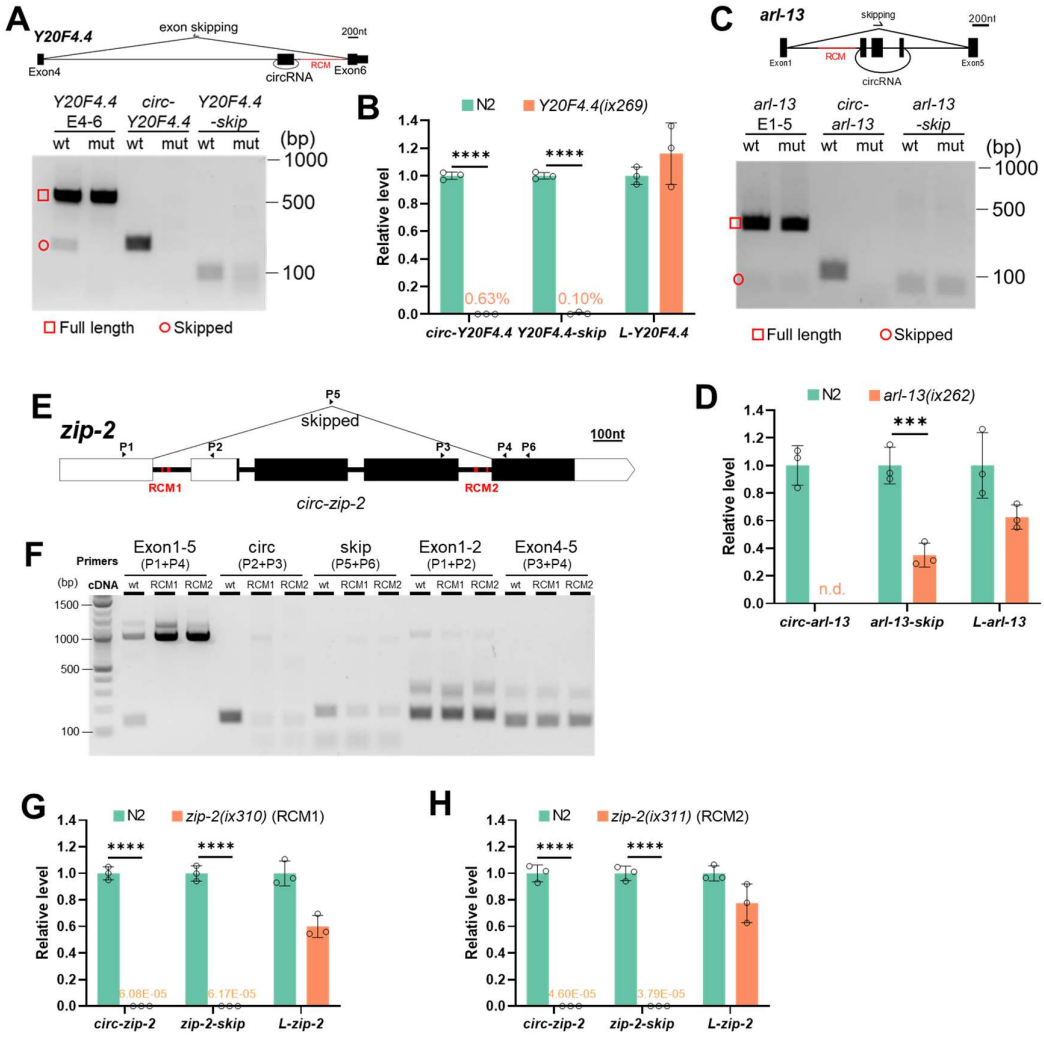


Figure 3.9 RCMs promote both back-splicing and exon-skipping. Modified from (*148*). (A) RT-PCR detection of *Y20F4.4* transcripts in wild-type N2 strain (wt) and the RCM-deleted *Y20F4.4(ix269)* strain (mut). Positions of deleted RCM sequences are in red. (B) RT-PCR detection of *arl-13* transcripts in wild-type N2 strain (wt) and the RCM-deleted *arl-13(ix262)* strain (mut). (C) RT-qPCR quantification of *Y20F4.4* transcripts in wild-type N2 strain and *Y20F4.4(ix269)* strain. (D) RT-qPCR quantification of *arl-13* transcripts in wild-type N2 strain and *arl-13(ix262)* strain. (E) Illustration of the gene structure of *zip-2*. P1-P6: positions of primers. Black rectangles indicate coding regions and white parts are untranslated regions (UTRs). RCM areas are in red. (F) RT-PCR detection of transcripts from *zip-2* gene in wild-type N2 strain and RCM-deleted strains. RCM1: *zip-2(ix310)*, RCM2: *zip-2(ix311)*. (G, H) RT-qPCR quantification of *zip-2* transcripts in RCM-deleted strains (*zip-2(ix310)* and *zip-2(ix311)*) compared with wildtype N2 strain. (C, D, G, H) Results are normalized to levels in N2 strain using *pmp-3* as the reference gene. Error bars are the standard deviations of three biological replicates. ***, $p < 0.001$, ****, $p < 0.0001$, two-tail Student's *t*-test. For all the RT-PCR and RT-qPCR results, RNA samples were from the L1 stage of the indicated strains.

3.2.6 RCM sequences in *zip-2* are highly conserved across several nematode species

Competing RNA secondary structures formed by base-pairing between introns that regulate mutually exclusive splicing are evolutionally conserved (*154, 155*). One study has also shown that conserved complementary sequences in introns are associated with exon-skipping (*156*). I then checked whether RCM sequences in *zip-2* are conserved or not. Ortholog genes of *zip-2* exist in five nematode species (*C. elegans*, *C. brenneri*, *C. briggsae*, *C. japonica,* and *C. remanei*). These *zip-2* genes have similar gene structures (Figure 5.13). Sequences in the upstream introns and downstream introns of these *zip-2* genes were aligned. Of the two pairs of RCMs in *zip-2* of *C. elegans*, the 13-nt RCMs are highly conserved across the five nematode species, both in the upstream introns and the downstream introns (Figure 3.10A and B). From available splicing data on WormBase, transcripts that skip exons bracketed by the conserved RCMs were found in all these *zip-2* genes (Figure 5.13, red arrows), suggesting the conserved RCMs possibly promote the conserved exon-skipping in all these *zip-2* genes.

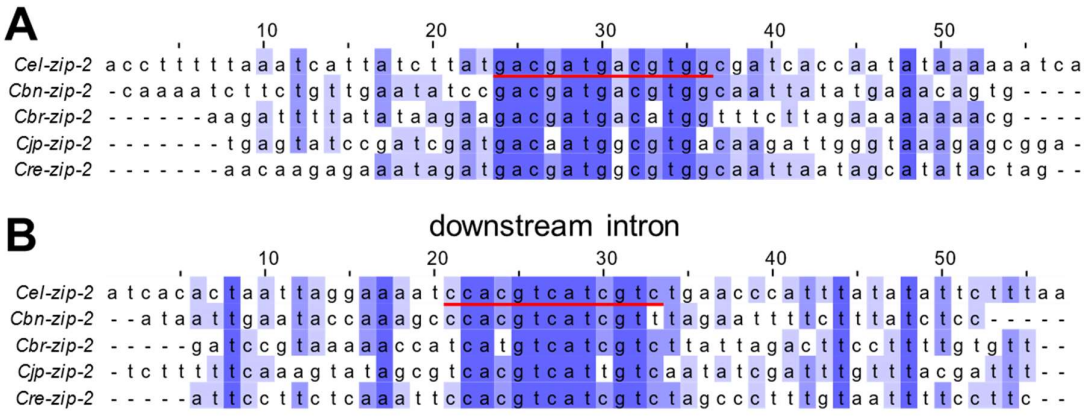


Figure 3.10 The 13-nt RCM sequences in *zip-2* are highly conserved in five nematode species. Modified from (*148*). Alignment of upstream (A) and downstream (B) intronic sequences in ortholog *zip-2* genes in indicated nematode species. Red lines underline the 13-nt RCM sequences in *zip-2* of *C. elegans*.

3.2.7 RCMs do not promote exon-skipping through back-splicing, neither the other way

Current knowledge suggests that RCMs promote circRNA formation by bringing the splicing sites for back-splicing in close proximity. Since the correlated back-splicing and exon-skipping use the same pair of introns, it is possible that RCMs also bring the splice sites for exon-skipping together. In principle, the y-shaped intermediate of back-splicing could be further spliced to form the corresponding skipped transcripts. Moreover, a previous study has shown that circRNA can be produced through a lariat intermediate produced by exon-skipping (*36*).

Whether RCMs promote exon-skipping first or back-splicing first? There are three possibilities: 1). RCMs promote back-splicing first; 2). RCMs promote exon-skipping first; 3). RCMs promote both back-splicing and exon-skipping directly at the same time (Figure 3.12). In order to clarify the three possibilities, the four splice sites (ss) and two branch points (BP) in intron 1 and intron 4 of *zip-2* were mutated one by one by CRISPR-Cas9 (Table 5.8 and Figure 5.14). The 5'ss in intron 1, BP, and 3'ss in intron 4 are used for exon-skipping; hence these sites are named skip-5'ss, skip-BP, and skip-3'ss, respectively. Similarly, BP and 3'ss in intron 1 and 5'ss in intron 4 are named circ-BP, circ-3'ss, and circ-5'ss, respectively (Figure 3.11A). For ss mutation, the conserved AG or GT nucleotides were deleted, and some possible cryptic splice sites nearby were mutated (Figure 5.15A and B). For BP mutation, since there is little information about BP sites in *C. elegans* (*157*), all A nucleotides were changed to G nucleotides in the 3' half of intron 1 and intron 4, without disturbing the RCM sequences. (Figure 5.15A and B).

Then, *zip-2* transcripts were detected by RT-PCR using different primer sets (Figure 3.11B and Figure 5.16B). Mutation of ss and BP for exon-skipping (skip-5'ss, skip-3'ss, or skip-BP) sufficiently abolished *zip-2-skip*. However, *circ-zip-2* was still produced in these strains (Figure 3.11B). For mutations of ss/BP required for back-splicing, circ-3'ss mutation produced a circRNA using a noncanonical AA site (*158*) (Figure 3.11B, Figure 5.15B, and Figure 5.16A). circ-5'ss and circ-BP mutation both blocked circRNA formation, but the skipped product can still be detected (Figure 3.11B). These results suggest that in *zip-2*, exon-skipping is not absolutely required for back-splicing and *vice versa*. RCMs can promote both exon-skipping and back-splicing directly at the same time.

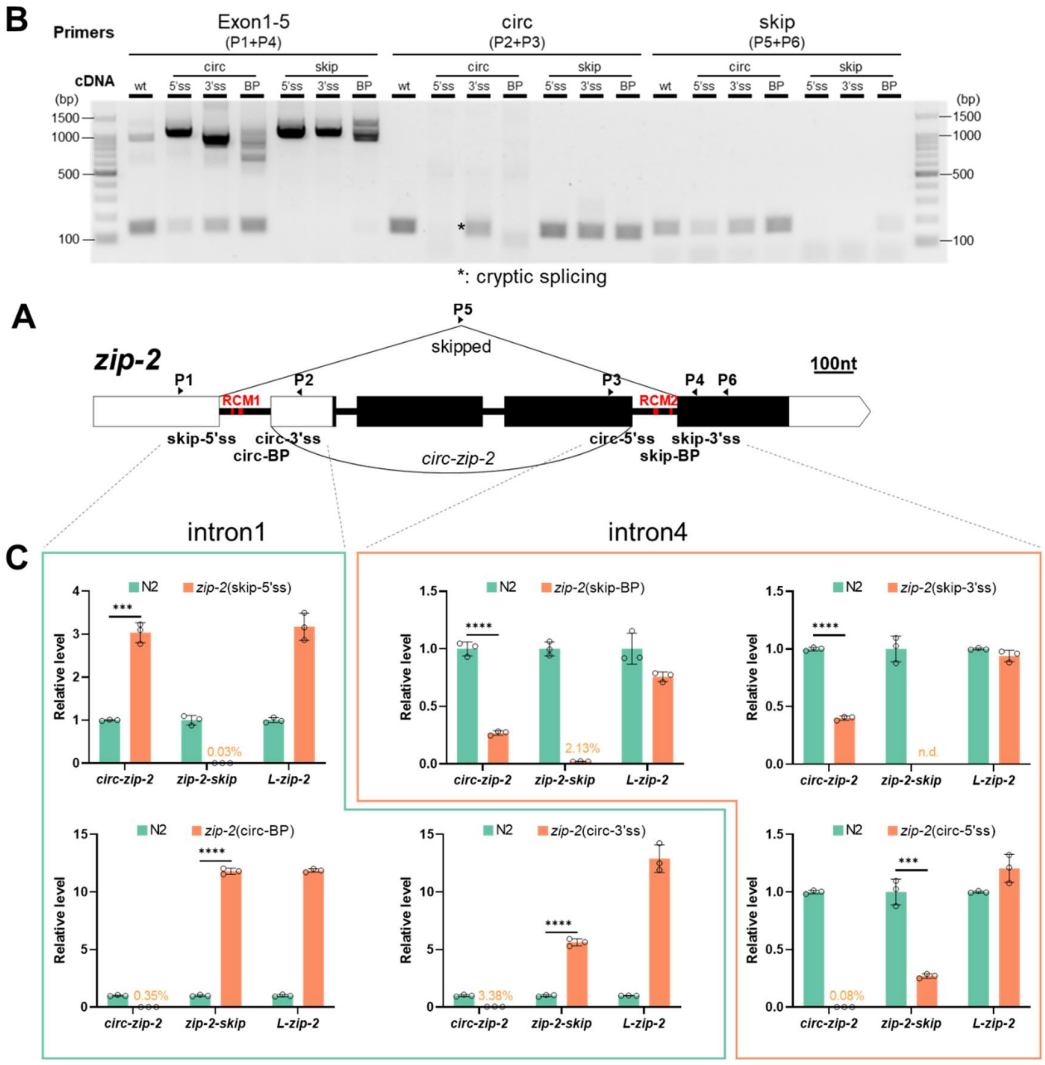


Figure 3.11 Detection and quantification of *zip-2* transcripts in ss/BP mutated strains. Modified from (*148*). (A) Gene structure of *zip-2*. P1-P6: positions of primers. Positions of splicing sites and branch points required for back-splicing and exon-skipping are labeled. Positions of RCMs are in red. (B) RT-PCR detection of *zip-2* transcripts in wild-type N2 strain and strains with mutated ss or BP. Note the cryptic splicing in the circ-3'ss strain. RNA samples from L1 worms were used for reverse transcription. (C) RT-qPCR quantification of levels *zip-2* circRNA (*circ-zip-2*), the skipped transcript (*zip-2-skip*), and the full-length linear transcript (*L-zip-2*) in the L1 stage of the indicated ss mutation strains and BP mutation strains. Results are normalized to levels in N2 strain using *pmp-3* as reference gene. Error bars are the standard deviations of three biological replicates. n.d.: not detected. ***, $p < 0.001$, ****, $p < 0.0001$, Two-tail Student's *t*-test.

Next, the levels of *circ-zip-2*, *zip-2-skip*, and *L-zip-2* in these mutant strains were quantified by RT-qPCR (Figure 3.11C). For *L-zip-2* quantification, primer set (P1 +P2) was used when mutated sites are in intron 4, and primer set (P3 + P4) was used when mutated sites are in intron 1. As expected, *zip-2-skip* was either undetectable or reduced to very low levels in strains with skip-5'ss, skip-3'ss, or skip-BP mutated. However, *circ-zip-2* in these three strains was either increased or decreased. Similar things were observed in strains with circ-BP, circ-3'ss or circ-5'ss mutated. Although *circ-zip-2* could be produced using the cryptic AA site in the circ-3'ss strain, the efficiency was much lower (Figure 3.11C, lower middle). *circ-zip-2* was sufficiently blocked in all these three strains. However, zip-*2-skip* showed increased levels in two strains (circ-BP and circ-3'ss) and a decreased level in the circ-5'ss mutated strain.

Interestingly, when the mutated sites were in the upstream intron (intron 1: skip-5'ss, circ-BP, and circ-3'ss), the transcripts that do not use the mutated sites, either *circ-zip-2* or *zip-2-skip*, were all increased (Figure 3.11C, intron 1). Also, the full-length linear mRNA was all increased in these mutants (Figure 3.11C, intron 1). Linear mRNA's increase may be due to enhanced transcription or increased stability after mutations in intron 1. If the transcription of *zip-2* is somehow enhanced by mutation of ss/BP in intron 1, it is reasonable that all the transcripts of *zip-2* get increased. If the increase of *L-zip-2* is not due to transcription enhancement, increased *circ-zip-2* or *zip-2-skip* in these mutated strains may be due to the competition of back-splicing and exon-skipping with the canonical splicing of intron 1. Since splicing is always happening during transcription (*159*), mutation of ss or BP in intron 1, which results in the intron retention or low-efficient splicing (Figure 5.14B), could increase the possibility to use the splicing sites in downstream introns, which could be either exon-skipping or back-splicing.

When the mutated sites were in the downstream intron (intron 4: circ-5'ss, skip-BP, and skip-3'ss), the transcripts that did not use the mutated sites were all decreased (Figure 3.11C, intron 4). Levels of full-length linear mRNA of *zip-2* did not change so much in these three strains. These results suggest that *circ-zip-2* could be produced from the exon-skipping pathway through the lariat intermediate (Figure 3.12). Also, *zip-2-skip* could be formed from the y-shaped intermediate of the back-splicing pathway (Figure 3.12).

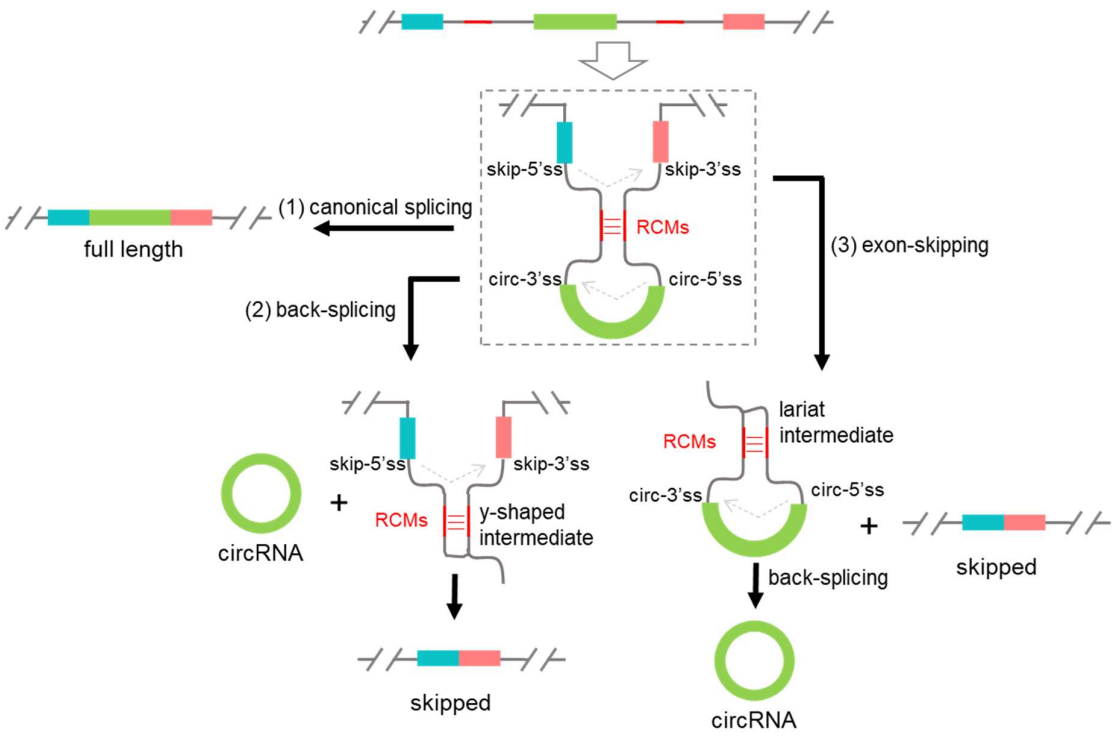


Figure 3.12 A proposed model that RCMs directly promote both back-splicing and exon-skipping at the same time. Adopted from (*148*). (1) Canonical splicing to form full-length linear mRNA. (2) RCMs facilitate circRNA formation by bringing splice sites for back-splicing sites together. The y-shaped intermediate is further spliced to form the skipped transcript. (3) RCMs promote exon-skipping by bringing splice sites for exon-skipping together. The lariat intermediate is further back-spliced to form circRNA. RCMs in the y-shaped intermediate and the lariat intermediate may help the second splicing steps.

3.3 circRNA regulation by RNA-binding protein FUST-1

Previous studies have shown that circRNAs are expressed in a tissue-specific and well-regulated manner (*52, 66, 68, 78, 127, 160*), suggesting the existence of specific factors that regulate circRNA production. Here, taking advantage of the available RBP mutants in *C. elegans*, I aimed to identify potential circRNA regulators.

Of note, except for results in Chapter 3.3.5 and 3.3.6, all results in Chapter 3.3 were from the author's published paper with modifications (*161*).

3.3.1 RBP screening identifies FUST-1 as a circRNA regulator

Several previously identified circRNAs that were either neuron-enriched or highly expressed in neurons were selected as targets (Figure 3.3D) (*148*). Thirteen RBPs that are conserved and have expressions in the neurons were chosen as potential regulators (*162*). Using mutant strains of these RBPs, a screening by RT-qPCR was performed to check the level changes of selected circRNAs in these mutant strains compared with wild-type N2 strain at the L1 stage (Figure 3.13A). As expected, levels of some circRNAs were altered in these mutant strains. Interestingly, most level changes of the selected circRNAs in these mutants were downregulations, suggestive of these RBPs' beneficial roles in circRNA production. Moreover, multiple neuron-enriched circRNAs (*circ-glr-2*, *circ-iglr-3*, *circ-arl-13*, *circ-cam-1*) were found to be downregulated in several strains (*asd-1(csb32)*, *tiar-3(csb35)*, *fox-1(csb39)*, *mec-8(csb22)*, *hrpf-1(csb26)*, and *fust-1(csb21)*), suggesting the regulation of these circRNAs by multiple RBPs. This is consistent with their roles in alternative splicing, where combinational regulation of one target by multiple RBPs is common in *C. elegans* (*163*). In line with this, no additive effect in circRNA regulation was found in *fust-1(csb21); hrpf-1(csb26)* double mutant strain compared with *fust-1(csb21)* single mutation (Figure 5.17A), suggesting that these RBPs may function as parts of a whole RNA-protein complex.

In these strains, *fust-1(csb21)* showed the most substantial downregulation of multiple circRNAs (Figure 3.13A). Hence it was chosen for further investigation. The *fust-1(csb21)* strain was made by replacement of *fust-1* genomic sequences with a *pharynx::GFP::NeoR* fragment by CRISPR-Cas9 (*162*) (Figure 3.13B). The downregulation of these circRNAs was also found in another *fust-1* mutant strain *fust-1*(*tm4439)*, which has a 240-bp deletion in intron 2 and exon 3 (Figure 3.13B and C). To further confirm the role of *fust-1* in circRNA regulation, a rescue strain (*fust-1(csb21); Ex[fust-1::mRFP]*) and an overexpression strain (*Ex[fust-1::mRFP]*) were made with extrachromosomal expression of the *fust-1* genomic sequences, starting from the promoter (2181 bp upstream ATG) to just before the stop codon. Monomeric red fluorescent protein (mRFP) was fused to the C-terminal to check expression patterns. The expression of FUST-1 was mainly in the nuclei of neurons and intestinal cells (Figure 3.13D, Figure 5.17B and C). The mRFP-positive L1 worms from the rescue strain and the overexpression strain were sorted, and levels of the circRNAs were checked by RT-qPCR. As expected, the levels of downregulated circRNAs were restored in the rescue strain, confirming *fust-1*'s role in promoting circRNA production (Figure 3.13E). The *fust-1(csb21)* strain also showed another phenotype of lower average moving speed at day three adult stage when cultured at 25ºC, which was also recovered in the rescue strain (Figure 3.13F). Although multiple copies of *fust-1* existed in the extrachromosomal arrays of the rescue and the overexpression strain (Figure 5.17D), these strains did not show much further improvement in circRNA levels or improvement in locomotion speed (Figure 3.13E and F). This may be because of post-transcriptional regulation of *fust-1* or saturation of FUST-1 protein.

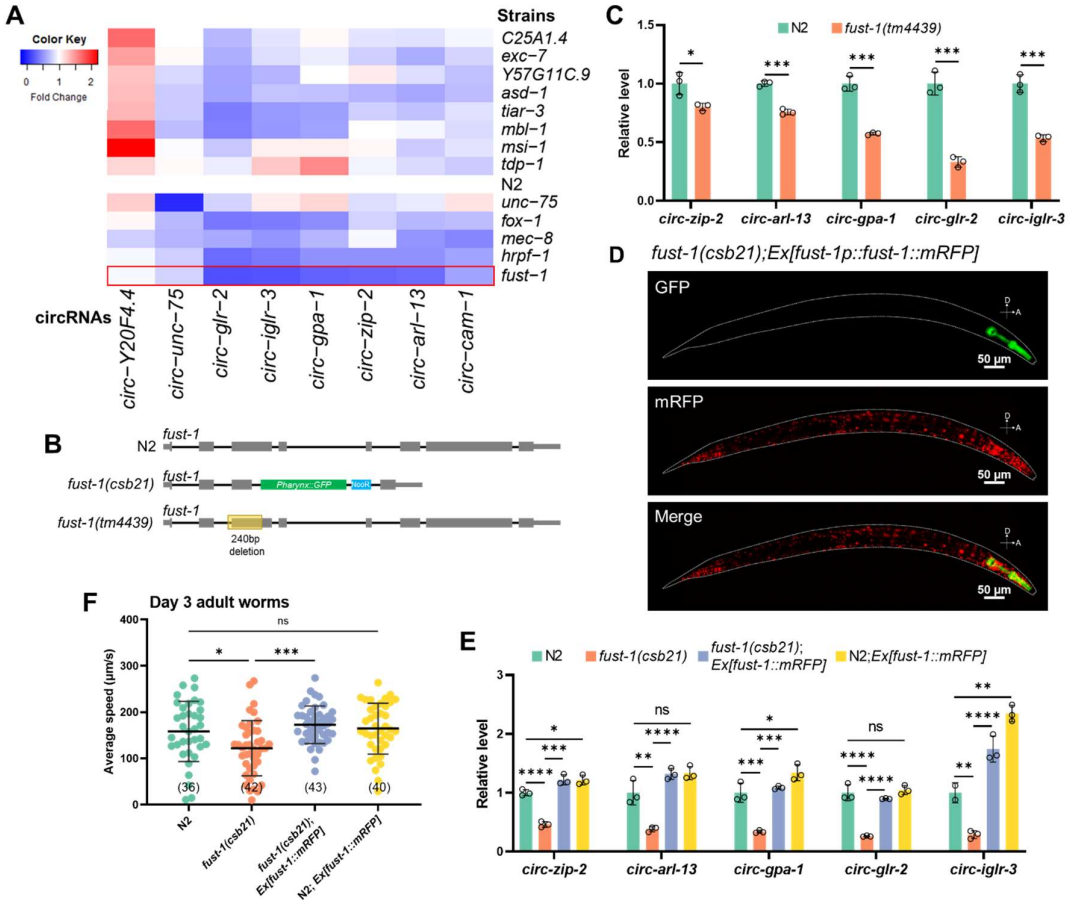


Figure 3.13 RBP screening identifies FUST-1 as a circRNA regulator. Modified from (*161*). (A) Heatplot showing the fold changes of circRNAs in 13 RBP mutant strains compared with wild-type N2 strain. Foldchanges are quantified by RT-qPCR and normalized to the N2 strain using *pmp-3* as the reference gene. RNA samples were from L1 worms of the indicated strains. Blue color means downregulation and red color means upregulation. (B) Gene structure of *fust-1* in wildtype N2 strain, *fust-1(csb21)*, and *fust-1(tm4439)* strain. (C) RT-qPCR quantification of circRNA levels at the L1 stage of wild-type N2 strain and *fust-1(tm4439)* strain. Levels are normalized to the N2 strain using *pmp-3* as the reference gene. Results are shown as mean ± sd of three biological replicates. Two-tailed Student's *t*-test. $p<0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. (D) Representative images showing the expression pattern of mRFP-fused FUST-1 in *fust-1(csb21)* strain. Worm stage: day 1 adult. Note the pharyngeal GFP expression in *fust-1(csb21)*. Scale bars: 50 μm. (E) RT-qPCR quantification of circRNAs at the L1 stage of the indicated strains. Levels are normalized to the N2 strain using *pmp-3* as the reference gene. Results are shown as mean ± sd of three biological replicates. (F) Average moving speed of day 3 adult worms raised at 25°C with the indicated genotypes. Numbers in brackets are numbers of worms used for moving speed measurement. Results are shown as mean ± sd. (E, F) One-way ANOVA, Tukey's multiple comparisons. $^{*}p<0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$; ns, not significant.

3.3.2 FUST-1 regulates circRNAs without affecting the cognate linear mRNAs

Next, to clarify whether FUST-1 promotes circRNA production by transcription promotion or not, levels of circRNAs and their cognate linear mRNAs were compared between the N2 strain and *fust-1(csb21)* strain at the L1 stage. While levels of these circRNAs were downregulated, their linear mRNA levels were not affected by the loss of FUST-1 (Figure 3.14A), indicating that FUST-1's role in circRNA production is not through promoting transcription. Using probes that hybridize to the circularized exons, which can detect both full linear and circular transcripts, I quantified *L-zip-2* and *circ-zip-2* between *fust-1(csb21)* and N2 strains by northern blot (Figure 3.14B and C). The results showed that while *circ-zip-2* was ~50% downregulated, *L-zip-2* was not slightly affected (Figure 3.14B and C).

To check the regulation of circRNA by FUST-1 globally, I performed RNA sequencing (RNA-seq) with ribosomal RNA depletion to compare differentially expressed circRNAs between *fust-1(csb21)* strain and wild-type N2 strain at the L1 stage. Similarly, circRNA annotation was performed using DCC (*131*), and differential expression was performed using DESeq2 (*133*). Both mRNAs and circRNA clustered separately in the PCA plots (Figure 5.18A and B). In total, 1266 circRNAs from 1199 annotated genes and 20 not-annotated loci were annotated with at least three BSJ reads in either group, with 916 in N2 strain and 849 in *fust-1(csb21)* strain (Figure 3.14D). TPM values of circRNAs were compared between the two strains. circRNAs in the N2 strain showed significantly higher TPM values than those in *fust-1(csb21)* strain (paired Wilcoxon test, $p < 2.2e-16$) (Figure 3.14E), indicating general promotional roles of FUST-1 in circRNA production, although some circRNAs were upregulated without FUST-1 (Figure 3.14F). Then, to check whether level changes in circRNA correlate with their cognate linear mRNAs, the fold changes of circRNA were plotted against those of their cognate mRNAs . The results showed a weak correlation (Pearson's correlation coefficient $R = 0.27$, $p = 3.5e-05$) of circRNAs with baseMean bigger than 3 (Figure 3.14F). The correlation was even weaker when all the circRNAs from annotated genes were considered (Figure 5.18C, Pearson's correlation coefficient $R = 0.14$, $p = 7.4e-07$). These results were consistent with the finding that FUST-1 regulates circRNAs without disturbing the cognate linear mRNA levels (Figure 3.14A).

Next, I asked whether FUST-1 has preferences in the regulation of neuronal circRNAs. The circRNAs identified in the "N2-*fust-1(csb21)*" dataset were compared with the previous "sort & whole" dataset (section 3.1), which gave 726 overlapped circRNAs (Figure 5.18D). Fold changes of the 726 overlapped circRNAs between N2 and *fust-1(csb21)* were plotted against those between the sort group (sorted neuron samples) and the whole group (whole worm samples). The results showed no correlation (Figure 3.14G, Pearson's correlation coefficient $R = -0.038$, $p = 0.3$), suggesting that FUST-1 has no preferences for neuronal circRNAs.

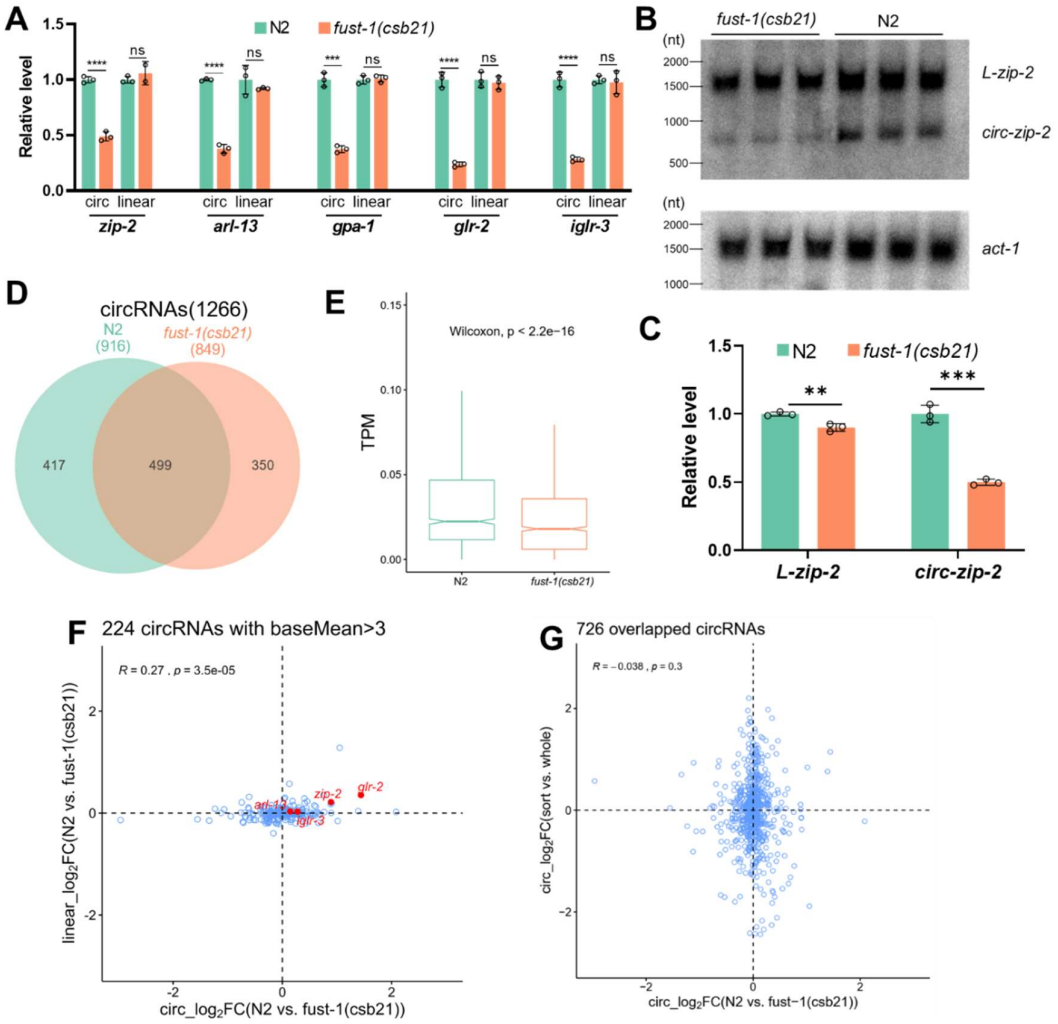


Figure 3.14 FUST-1 regulates circRNAs without affecting the cognate linear mRNAs. Modified from (*161*). (A) RT-qPCR quantification of circRNAs and their linear mRNAs at the L1 stage in the N2 strain and *fust-1(csb21)* strain. Levels are normalized to the N2 strain using *pmp-3* as the reference gene. Results are shown as mean ± sd of three biological replicates. Two-tailed Student's *t*-test. ***$p < 0.001$, ****$p < 0.0001$; ns, not significant. (B) Northern blot detection *zip-2* transcripts, both linear and circular, and *act-1* in the L1 stage of N2 strain and *fust-1(csb21)* strain. Total RNA samples (5 µg per lane) of 3 biologicla replicates from L1 worms were used. (C) Quantification of northern blot results in (B), normalized to N2 strain using *act-1* as the reference gene. Results are shown as mean ± sd. Two-tailed Student's *t*-test. **$p < 0.01$, *** $p < 0.001$. (D) Overlap of circRNAs detected in the RNA-seq results of N2 strain and *fust-1(csb21)* strain. (E) TPM (transcripts per million reads) comparison of all circRNAs between N2 and *fust-1(csb21). P* value indicates paired Wilcoxon test. (F) Scatter plot showing the log2 fold changes of 224 circRNAs with baseMean > 3 versus log2 fold changes of their corresponding linear mRNAs. The Pearson correlation coefficient (*R*) and *p* value (*p*) are shown. Names of several circRNA genes are labeled. (G) Scatter plot showing the log2 fold changes of 726 overlapped circRNAs between the "N2-*fust-1(csb21)*" dataset and the "sort-whole" dataset. The Pearson correlation coefficient (*R*) and *p* value (*p*) are shown.

3.3.3 FUST-1 binds to pre-mRNAs of circRNA genes

FUS binds to flanking introns of circRNA genes in N2a cells (*61*). I next checked whether FUST-1 in *C. elegans* recognizes pre-mRNAs of circRNA genes to regulate circRNA formation. Using CRISPR-Cas9 technology, I inserted the FLAG tag to the N terminal, just after the start codon, or to the C-terminal, just before the stop codon, respectively (Figure 3.15A and Figure 5.19A). The effect of FLAG-tag insertion on FUST-1's role in circRNA regulation was evaluated. While N-terminal FLAG insertion showed slight increases in circRNA levels, C-terminal FLAG tag fusion affected FUST-1's function in circRNA regulation in multiple circRNAs (Figure 5.19B and C). Hence N-terminal FLAG fused FUST-1 strain was used for the co-immunoprecipitation (Co-IP) experiment. Dynabeads Protein G conjugated with anti-FLAG antibody (+Ab) were used for Co-IP. Beads only (-Ab) were used as the negative control. As expected, the anti-FLAG antibody successfully enriched FLAG::FUST-1 after Co-IP (Figure 3.15B and Figure 5.19D). Then the levels of pre-mRNAs of circRNA genes were quantified by RT-qPCR. Threshold cycle (Ct) values were used for comparison. Lower Ct values indicate higher levels. While rRNA control (*18S rRNA* and *26S rRNA*) was depleted after Co-IP, all the pre-mRNAs of circRNA genes were enriched compared with input samples (Figure 3.15C). Moreover, these pre-mRNAs showed significate lower Ct values than those of control groups without using of antibody (Figure 3.15C), suggesting that FUST-1 binds to pre-mRNAs of the circRNAs genes to regulate circRNA formation.

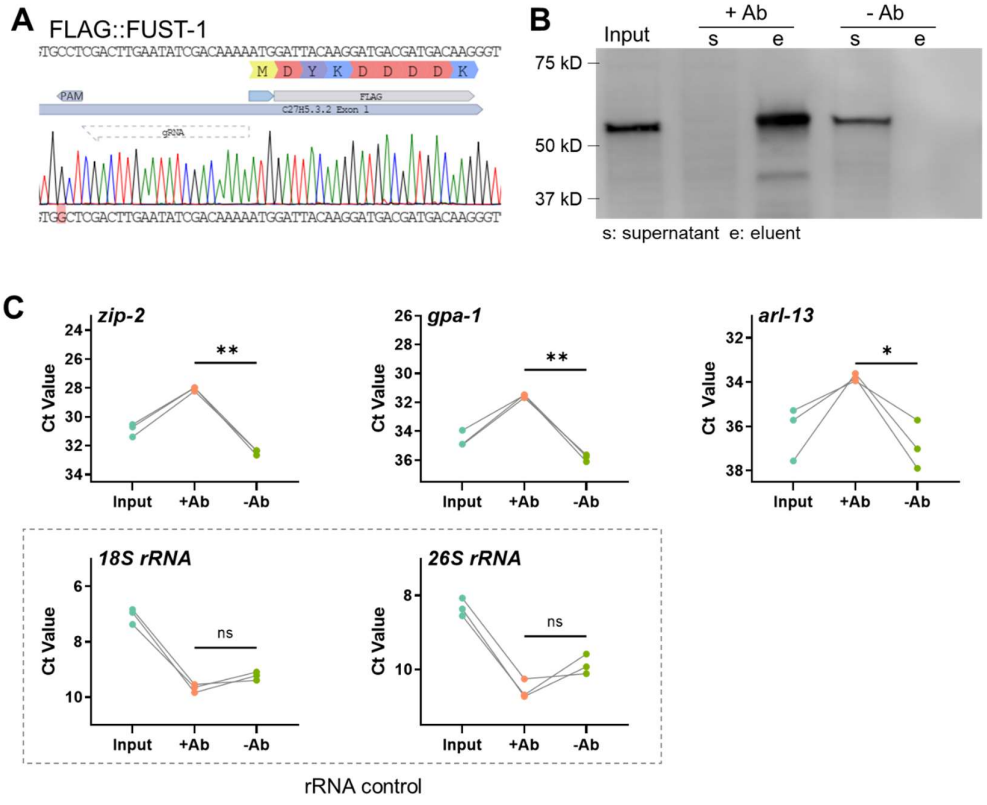


Figure 3.15. FUST-1 binds to pre-mRNAs of circRNA genes. Modified from (*161*). (A) Sequence confirmation for N-terminal fusion of the FLAG tag just after the start codon of FUST-1. Note the position of gRNA and the mutated PAM site (AGG>AGC). (B) Western blot showing the co-immunoprecipitation (Co-IP) of FLAG::FUST-1. (C) Ct value changes of pre-mRNAs of some circRNA genes before and after Co-IP of FLAG::FUST-1 with or without anti-FLAG antibody. Results from 3 biological replicates are shown. Paired two-tailed Student's *t*-test. *$p < 0.05$, **$p < 0.01$, ns, not significant.

3.3.4 FUST-1 regulates both exon-skipping and back-splicing

In section 3.2.4, transcripts that skip the exons to be circularized were identified in several circRNA genes. As the homolog of FUST-1 in humans and mice, FUS is involved in the regulation of AS of many genes by binding to their pre-mRNAs (*164-166*). Since FUST-1 binds to the pre-mRNAs of these circRNA genes, I then checked whether FUST-1 regulates exon-skipping or not. In *zip-2*, reads aligned to the skipped junction were much less in *fust-1(csb21)* strain (27.0 reads on average) than those in wild-type N2 strain (71.3 reads on average) (Figure 3.16A). The RT-qPCR quantification results also showed that both the circRNA and the skipped transcript in *zip-2* were downregulated without FUST-1 (Figure 3.16B). In *arl-13*, while the circRNA got downregulated in *fust-1(csb21)*, the skipped transcript was weakly upregulated (Figure 3.16C). These results suggest that FUST-1 may function differently in different genetic environments.

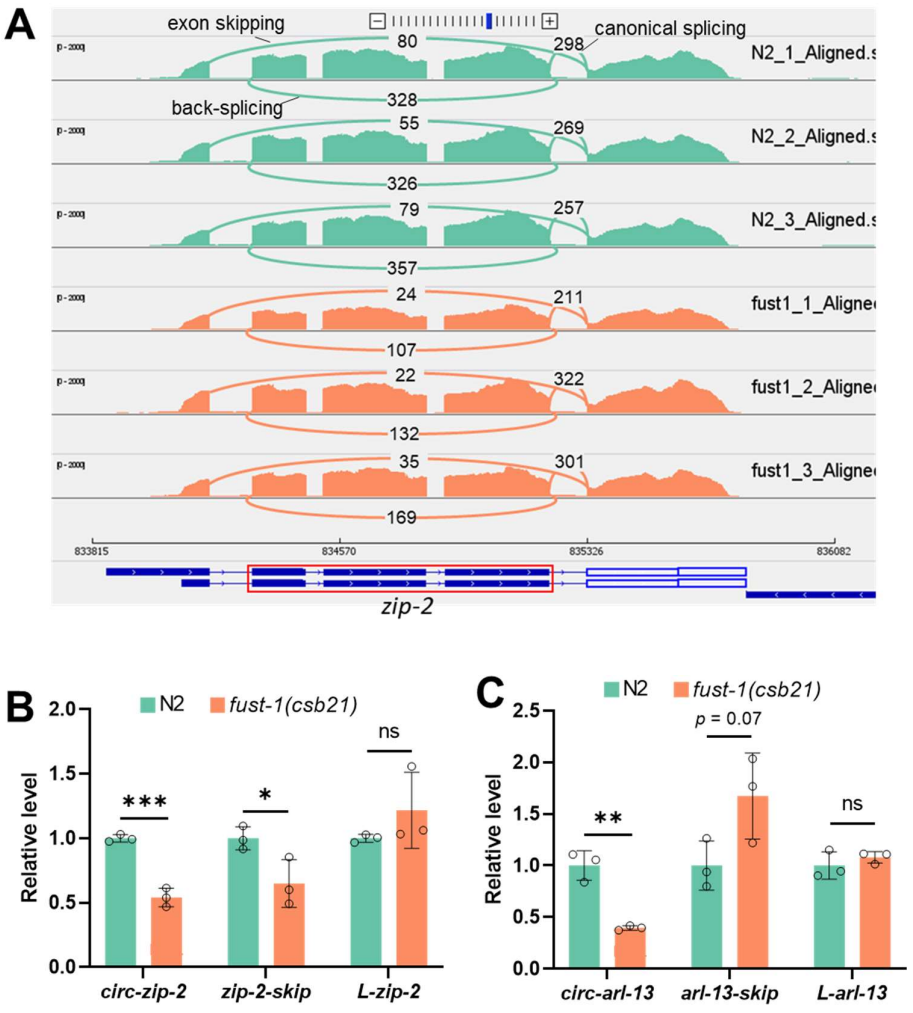


Figure 3.16 FUST-1 regulates both exon-skipping and back-splicing. Modified from (*161*). (A) Sashimi plot showing numbers of reads aligned to the canonical splice junction, the skipped junction, and the back-splice junction in *zip-2*. Exons in the red rectangle are circularized. (B, C) RT-qPCR quantification of levels of the circular, skipped, and full-length linear transcripts in *zip-2* (B) and *arl-13* (C) between wild-type N2 strain and *fust-1(csb21)* strain. RNA samples were from L1 worms. Levels are normalized to the N2 strain using *pmp-3* as the reference gene. Results are shown as mean ± sd of three biological replicates. Two-tailed Student's *t*-test. *$p < 0.05$, **$p < 0.01$, *** $p < 0.001$, ns, not significant.

3.3.5 The 5' splice site of exon-skipping and back-splicing are important for FUST-1's role in regulating back-splicing and exon-skipping, respectively

Next, I asked whether FUST-1 regulates exon-skipping and back-splicing independently or not. Taking advantage of the *zip-2* ss/BP mutant strains generated in section 3.2.7, I crossed them with *fust-1(csb21)* and obtained six ss/BP-*fust-1* double mutation strains.

Since mutation in circ-ss/BP and skip-ss/BP sufficiently blocked the production of *circ-zip-2* and *zip-2-skip*, respectively (Figure 3.11), I compared the levels of *zip-2-skip* between circ-ss/BP single mutation strains and the corresponding circ-ss/BP-*fust-1* double mutation strains. Similarly, levels of *circ-zip-2* were compared between skip-ss/BP and the corresponding skip-ss/BP-*fust-1* strains. In this way, we can tell whether FUST-1 still promotes back-splicing when exon-skipping is abolished and whether FUST-1 still promotes exon-skipping when back-splicing is blocked. Linear full-length *zip-2* mRNA was also quantified.

As shown in Figure 3.17B (middle and right panels), FUST-1 still promotes exon-skipping when circ-3'ss or circ-BP was mutated. However, when circ-5'ss was mutated, loss of FUST-1 has no effect on the level *of zip-2-skip* (Figure 3.17B, left panel), suggesting that the 5'ss of back-splicing (circ-5'ss) is important for FUST-1 to regulate exon-skipping. Similar results were found between skip-ss/BP and skip-ss/BP-*fust-1* comparisons: while *circ-zip-2* got downregulated in the absence of FUST-1 in the skip-3'ss or skip-BP background (Figure 3.17C, middle and right panels), its level did not change without FUST-1 when skip-5'ss was mutated (Figure 3.17C, left panel), suggesting that skip-5'ss is involved in FUST-1's role in back-splicing regulation.

Since 5'ss is involved in the interaction with U1 snRNP, I then asked whether FUST-1 interacts with components in U1 snRNP. Previous studies have shown that FUS in humans interacts with U1 snRNP (*167, 168*). Particularly, a recent paper showed that FUS in humans directly binds to stem-loop 3 of U1 snRNA and sequences in introns to regulate splicing (*169*). Hence, I checked whether FUST-1 could bind to U1 snRNA. As expected, U1 snRNA in *C. elegans* got significantly enriched after Co-IP of FLAG::FUST-1 (Figure 3.17D).

Intron 1 and intron 4 in *zip-2* are quite short, and secondary structures can be formed by base pairing between RCMs. Hence the spliceosome complexes formed for exon-skipping and back-splicing may be very close spatially. FUST-1 in the exon-skipping complex may interact with the back-splicing complex and promote back-splicing efficiency. Similarly, FUST-1 in the back-splicing complex may also interact with the exon-skipping complex and promote exon-skipping efficiency (Figure 3.17E).

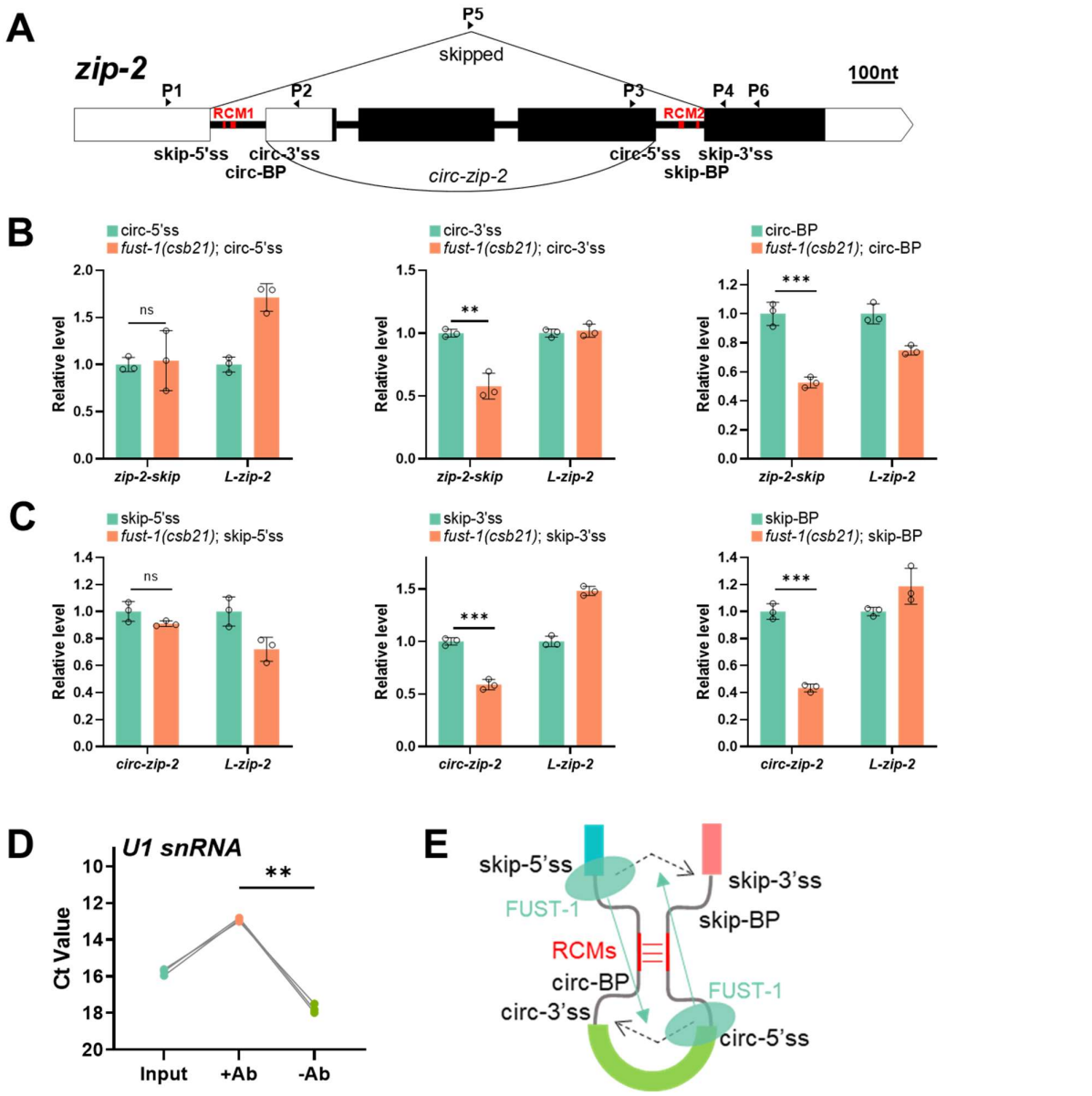


Figure 3.17 *zip-2* transcripts in ss/BP-*fust-1* double mutant strains compared with ss/BP single mutant strains. (A) Gene structure of the *zip-2* gene. (B, C) RT-qPCR quantification of *zip-2* transcripts between the ss/BP single mutation strains and the corresponding ss/BP-*fust-1* double mutation strains. RNA samples were from L1 worms. Levels are normalized to the ss/BP single mutation strains using *pmp-3* as the reference gene. Results are shown as mean ± sd of three biological replicates. Two-tailed Student's *t*-test. $^{**}p < 0.01$, $^{***}p < 0.001$, ns, not significant. (D) Ct value changes of *U1 snRNA* before and after Co-IP of FLAG::FUST-1 with or without anti-FLAG antibody. Results from 3 biological replicates are shown. Paired two-tailed Student's *t*-test. $^{**}p < 0.01$. (E) A model showing that FUST-1 binding to the skip-5'ss promotes back-splicing, and FUST-1 binding to circ-5'ss promotes exon-skipping.

3.3.6 FUST-1 knock-in mutants affect circRNA levels

Many naturally occurred mutations in FUS are found in the C-terminal nuclear localization signal (NLS) region, which cause cytoplasmic mislocalization and aggregation of FUS (*170, 171*). Several residues in the C-terminal of FUST-1 are conserved to those in the NLS of human FUS (Figure 3.18A). The R446S and P447L mutations in FUST-1 have been used to mimic R524S and P545L in FUS and result in disrupted autophagy and neuronal dysfunction in *C. elegans* (*172*). Here, I designed the same mutated sites in FUST-1 and tagged them with mRFP to check whether these mutations affect nuclear localization and circRNA production (Figure 3.18B). The insertion of mutated sites and mRFP was achieved by CRISPR-Cas9, and the sequences were confirmed (Figure 5.20A).

While R446S mutation did not obviously affect the nuclear localization of FUST-1, P447L mutated FUST-1 was found ubiquitously in the expressing cells (Figure 3.18C). At the egg stage, while wild-type FUST-1 and R446S FUST-1 were in the nucleus, P447L FUST-1 was absent from the nucleus, with expression only in the cytoplasm (Figure 5.20B).

Next, circRNA levels in these strains were compared. The mutations did not significantly affect *fust-1* mRNA levels (Figure 3.18D). Although R446S mutation did not change nuclear localization, all tested circRNA levels were affected (*circ-zip-2*, *circ-arl-13, circ-iglr-3, circ-gpa-1*) (Figure 3.18D), suggesting some other functions of the C-terminal of FUST-1. The P447L mutation also dramatically affected almost all the checked circRNA levels, with some circRNA (*circ-iglr-3* and *circ-gpa-1*) levels even lower than those of the *fust-1(csb21)* strain (Figure 3.18D), indicating nuclear localization of FUST-1 is critical for its regulation on circRNA.

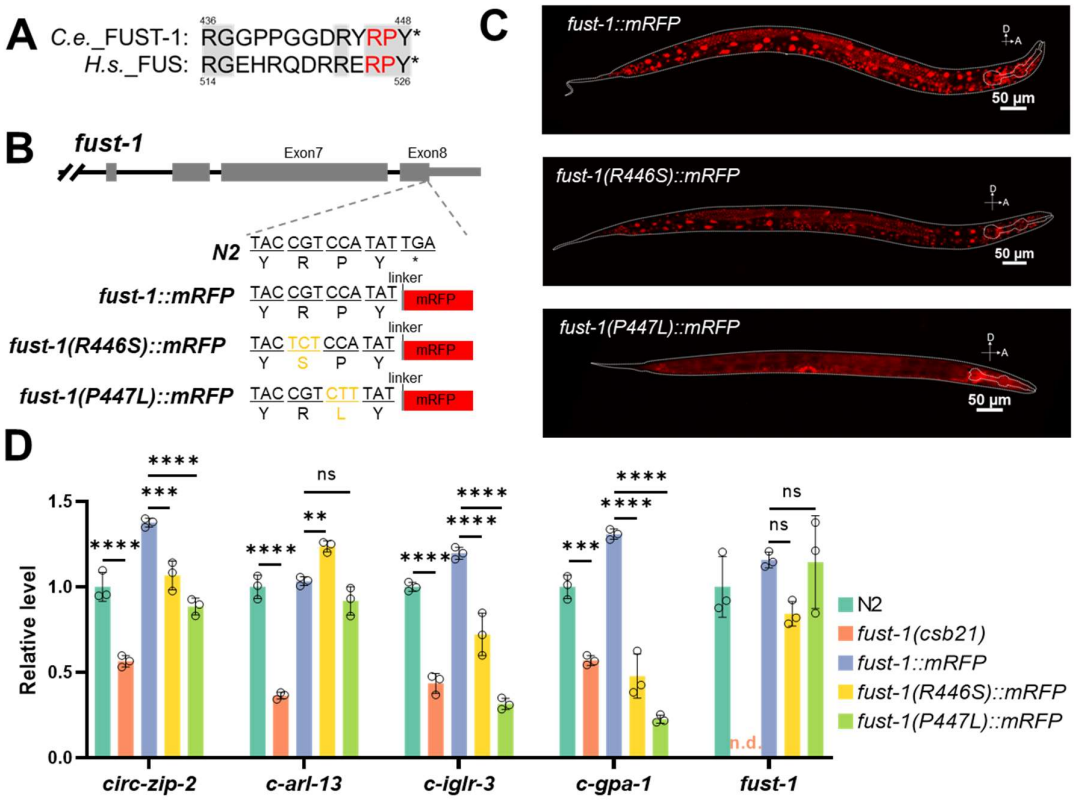


Figure 3.18. FUST-1 knock-in mutants affect circRNA levels. (A) C-terminal alignments of FUST-1 in *C. elegans* and FUS in humans. (B) Scheme showing the mutated sites and position of mRFP tagging. (C) Confocal images showing the expression patterns of indicated strains. Note the loss of nuclear localization of P447L mutated FUST-1. Worm stage: day 1 adult. A: Anterior, D: Dorsal. Scale bar: 50µm. (D) RT-qPCR results of circRNA levels in the indicated strains at the L1 stage. Levels are normalized to the N2 strain using *pmp-3* as the reference gene. Results are shown as mean ± sd of three biological replicates. n.d.: not detected. One-way ANOVA, Tukey's multiple comparisons. $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$; ns, not significant.

3.4 An autoregulation loop in *fust-1* for circRNA regulation

The overexpression of FUST-1 genomic sequences in the N2 strain did not show much further enhancement in either circRNA levels or the locomotion speed (Figure 3.13E and F). These results suggest that there may be post-transcriptional mechanisms to regulate FUST-1 protein functions. Of note, results in Chapter 3.4 were from the author's published paper with modifications (*161*).

3.4.1 An autoregulation loop in *fust-1*

FUST-1 protein has two isoforms: FUST-1, isoform a (FUST-1a) is from the full-length transcript, and FUST-1, isoform b (FUST-1b) is from the transcript with skipped exon 5 (Figure 3.19A). Moreover, FUST-1b is translated using a downstream AUG and a different reading frame (+1) compared with isoform a. The reading frame in FUST-1b becomes the same as in FUST-1a after the skipping of exon 5 (38 nt in length), which results in a shorter isoform with different N-terminal sequences. Still, the RNA recognition motif (RRM), zinc-finger (ZnF) domain, and the nuclear localization signal (NLS) domain in FUST-1b are the same as in FUST-1a (Figure 3.19A). To check how these two isoforms are expressed, two plasmids with different colors and a nonsense mutation in the reading frame of either isoform a (*fust-1a-mut::mRFP*) or isoform b (*fust-1b-mut::GFP*) were constructed so that only the other isoform can be expressed (Figure 3.19A and Figure 5.21A). Co-injection of the two plasmids in wild-type N2 strain showed that the two isoforms of FUST-1 were co-expressed in the nuclei of the same cells: neurons and intestinal cells (Figure 3.19B and Figure 5.21B). Interestingly, in early eggs, FUST-1a was expressed earlier than FUST-1b (Figure 3.19C). Furthermore, the *fust-1a-mut::GFP* plasmid was poorly expressed in *fust-1(csb21)* strain, and co-injection with *fust-1b-mut::mRFP* can increase the GFP intensity (observation during transgenic strain preparation). These results gave a hint that FUST-1a may promote the production of FUST-1b.

To prove this hypothesis, I constructed a dual-color splicing reporter (*173, 174*) of the skipping of exon 5 in *fust-1* with a neuronal promoter (Figure 5.21), in which no skipping gives GFP expression while skipping of exon 5 results in mCherry expression. As expected, two colors were co-expressed in almost all the neurons in the wild-type strain (Figure 3.19D), suggesting that exon-skipping of exon 5 is happening in all the neurons. However, when the reporter plasmid was crossed into two *fust-1* mutation strains, *fust-1(csb21)* and *fust-1(tm4439)* (Figure 3.13B), the expression of mCherry was dramatically reduced (Figure 3.19D and Figure 5.21D), indicating FUST-1 was involved in the exon-skipping of its own pre-mRNA. Since *fust-1(csb21)* strain has pharyngeal GFP expression (*162*) (Figure 3.13B and Figure 3.19D), neurons in the ventral nerve cord around the neck were used to quantify the mCherry-to-GFP intensity ratios (Figure 3.19D). The mCherry-to-GFP ratios were significantly reduced in both two *fust-1* mutants, and they did not change in the *mec-8(csb22)* strain (Figure 3.19E and Figure 5.21D).

Next, to prove that FUST-1a promotes the skipping of exon 5 of *fust-1* pre-mRNA, I tried the rescue of mCherry expression of the splicing reporter in *fust-1(csb21)* by co-injection of the reporter plasmid with FUST-1a cDNA or FUST-1b cDNA, driven by the *fust-1* original promoter (2181 bp upstream the ATG of FUST-1a). One more construct with truncated N-terminal (FUST-1-ΔN) was also used (Figure 5.22A). Tail-expressing plasmid *lin-44p::mRFP* was used as an injection marker. As expected, isoform a cDNA restored the mCherry expression of the splicing reporter, while isoform b cDNA did not (Figure 3.19F, Figure 5.22B and C), which confirms that FUST-1a promotes the skipping of exon-5 to

produce FUST-1b. Consistent with this, *fust-1* pre-mRNA, detected by primers in intron 4 of *fust-1*, was significantly enriched after Co-IP with FLAG::FUST-1, which only tagged FUST-1a (Figure 3.19G). To my surprise, the FUST-1-ΔN construct also rescued the mCherry expression, just as efficient as FUST-1a (Figure 3.19F and Figure 5.22D). Since the three isoforms have identical functional domains (RRM, ZnF, and NLS) with different N-terminal sequences (Figure 5.22A), these results suggest that the N-terminal sequences in isoform a may not be so crucial for its function, and the N-terminal in FUST-1b may prevent its domains from functioning normally.

Taken together, I characterized an autoregulation loop in *fust-1*, in which FUST-1a promotes the skipping of exon 5 of *fust-1* pre-mRNA, resulting in the production of FUST-1b.

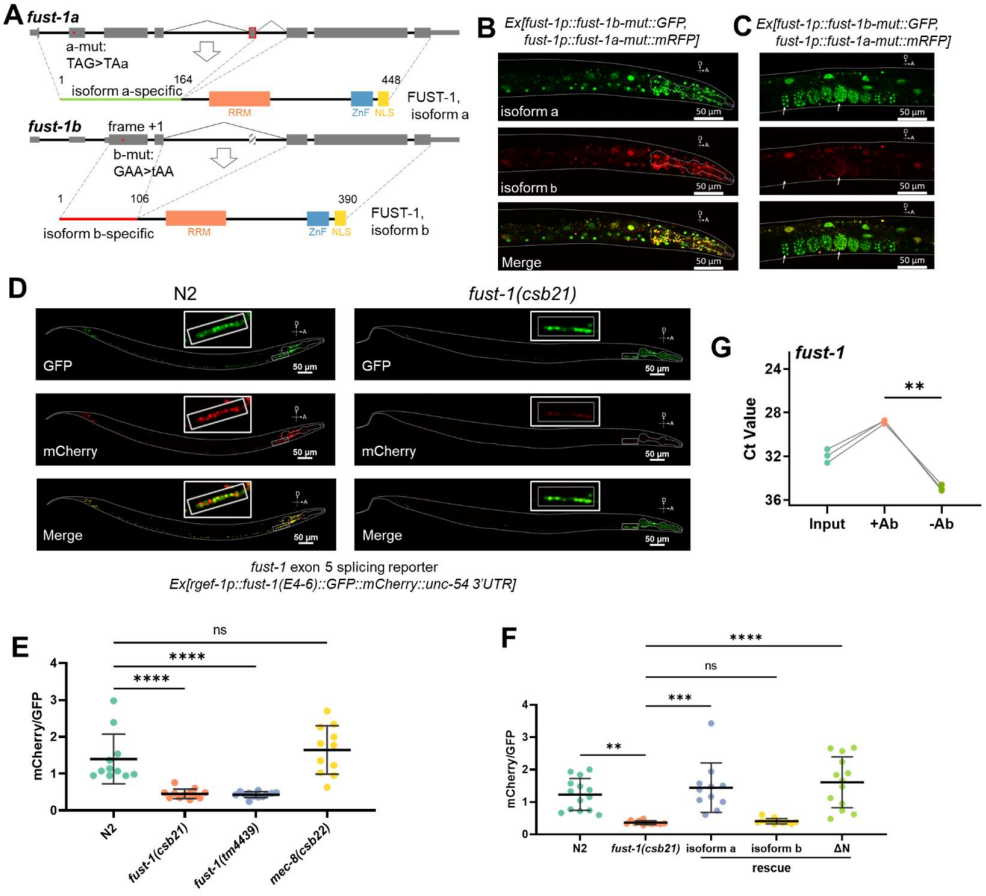


Figure 3.19. An autoregulation loop in *fust-1*. Modified from (*161*). (A) Gene structure of *fust-1* and the domains in FUST-1a and FUST-1b. Note the positions where nonsense mutations were introduced (Red asterisks). Lengths of amino acids in each isoform were labeled. RRM: RNA-recognition motif; ZnF: Zinc-figure; NLS: nuclear localization signal. (B, C) Confocal images showing expression of FUST-1a and FUST-1b in the nuclei of neuron cells (B) and eggs (C). Worm stage: day 1 adult. Note that in early eggs, FUST-1a was expressed earlier than FUST-1b (white arrows). A: Anterior, D: Dorsal. Scale bars: 50 μm. (D) Representative confocal images showing the expression patterns of splicing reporter of *fust-1* exon 5 in N2 strain and *fust-1(csb21)* strain. Inset squares show the enlarged neck neurons used for mCherry-to-GFP ratio quantification. Worm stage: day 1 adult. A: Anterior, D: Dorsal. Scale bars: 50 μm. (E, F) Quantification of mCherry-to-GFP ratios of the *fust-1* exon5 splicing reporter in the indicated strains. Young adult to day 1 adult worms were used. One-way ANOVA, Tukey's multiple comparisons. $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$; ns, not significant. (G) Ct value changes of *fust-1* pre-mRNA before and after Co-IP of FLAG::FUST-1 with or without anti-FLAG antibody. Primer positions are in intron 4 of *fust-1* pre-mRNA. Results from 3 biological replicates are shown. Paired two-tailed Student's *t*-test. $^{**}p < 0.01$.

3.4.2 FUST-1a is the functional isoform in circRNA regulation

Next, to check which isoform of FUST-1 is functional in circRNA regulation, I tried to rescue the downregulated circRNAs in *fust-1(csb21)* with extrachromosomal expression of FUST-1 isoform cDNA with C-terminal mRFP fusion, in which either FUST-1a or FUST-1b or FUST-1-ΔN is expressed (Figure 5.23). The mRFP-positive L1 worms were sorted, from which total RNA was extracted, and then circRNA levels were quantified by RT-qPCR. Same with their roles in exon-skipping, FUST-1a successfully rescued the downregulated circRNAs, whereas FUST-1b did not improve the downregulated circRNA levels at all, indicating that FUST-1a is the functional protein in circRNA regulation (Figure 3.20A). Although not as efficient as FUST-1a, FUST-1-ΔN fully rescued the downregulated *circ-zip-2* and *circ-iglr-3* and partially restored *circ-arl-13* level (Figure 3.20A).

In an effort to generate strains with C-terminal mRFP tagging of FUST-1 isoforms, I achieved C-terminal mRFP insertion in *fust-1* (*fust-1::mRFP*, section 3.3.6). Another obtained strain, in which intron 3 to intron 6 of *fust-1* were removed, cannot use the autoregulation pathway, resulting in the expression of only FUST-1a (*fust-1a::mRFP*) (Figure 3.20B). I failed to obtain a strain that can only express mRFP tagged FUST-1b. Consistent with the extrachromosomal expression pattern of FUST-1 (Figure 3.13D, Figure 5.17B and C), mRFP-tagged FUST-1 was mainly expressed in the nuclei of neurons and intestinal cells (Figure 3.20C). Moreover, FUST-1 was also found in the nuclei of gonads (Figure 3.20D), which was not observed in extrachromosomal expression, probably due to silencing of the multicopy transgenes in the germline (*175*).

Levels of circRNAs were compared between the two strains. Out of the five checked circRNA, the levels of four circRNAs were altered in the strain only FUST-1a can be expressed (Figure 3.20E), suggesting the autoregulation loop is critical for FUST-1's role in circRNA regulation.

Some results from chapter 3.3 and chapter 3.4 have been published with modifications (*161*).

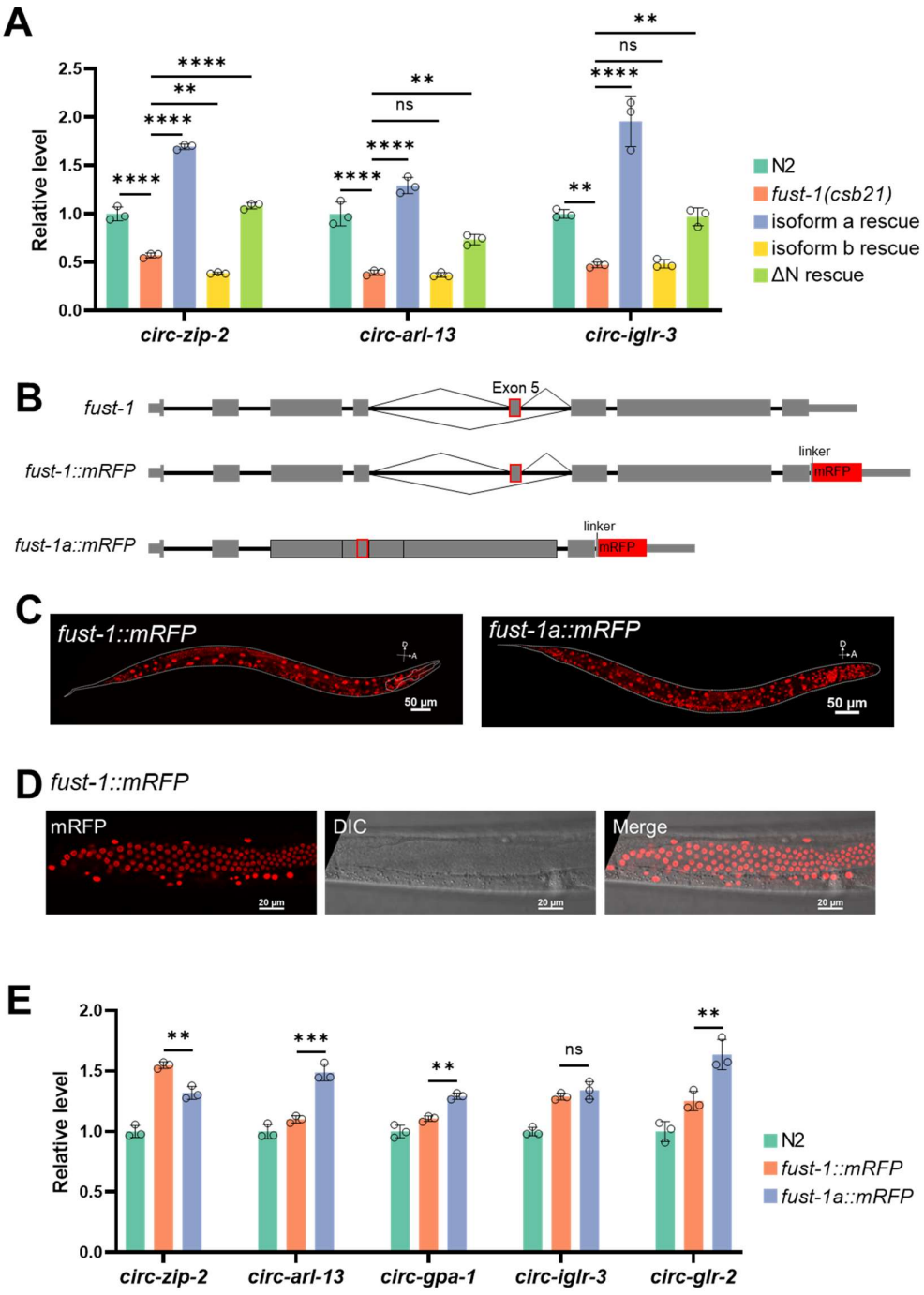


Figure 3.20 FUST-1a is the functional isoform in circRNA regulation. Modified from (*161*). (A) Rescue of circRNA levels by FUST-1 isoforms, quantified by RT-qPCR. cDNA samples from L1 worms of indicated strains were used. (B) Illustration of the gene structures of *fust-1* in wild-type N2 strain, C-terminal mRFP fused *fust-1* (*fust-1::mRFP*) strain, and the strain can only express FUST-1a (*fust-1a::mRFP*). (C) Representative images showing expression patterns of FUST-1 in the indicated strains. Worm stage: young adult. A: Anterior, D: Dorsal. Scale bars: 50 μm. (D) Confocal images showing that FUST-1 is expressed in the gonad. Scale bar: 20 μm. (E) RT-qPCR quantification of circRNA levels at the L1 stage of indicated strains. (A, E) Levels are normalized to N2 strain using *pmp-3* as the reference gene. Results are shown as mean ± sd of three biological replicates. One-way ANOVA, Tukey's multiple comparisons. $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$; ns, not significant.

4. Discussion

4.1 Large-scale neuron sorting

In this study, I optimized a method for large-scale neuron isolation from L1 worms. The amount of obtained RNA from sorted neurons was increased for the first time to hundreds of nanogram scale, making the detection of circRNA by RNA-seq more reliable. This method can be readily expanded to the sorting of other types of cells for cell- or tissue-specific transcriptome profiling. The "labeling-dissociation-sorting" method has been used to isolate specific cells from different developmental stages, from dissociated egg cells to adult cells (*130, 140-145*). One limitation of this method to isolate neuron cells is that neurites can get lost after dissociation and sorting. Unlike structured shapes of neuron cells in the worm body, most neurons show sphere shapes with little neurites after dissociation (Figure 3.1C and Figure 5.1B). The neurite structures may get broken during dissociation and sorting, or they may retract to the cell body after getting rid of their surrounding cells. In the former case, the RNA molecules in the neurite cannot be captured. Adding propidium iodide (PI) or other nuclear staining dyes can be used to label dead/broken cells, which can be helpful to exclude damaged neuron cells.

4.2 Neuronal circRNA profile

Using this method, I provided the first neuronal circRNA profiles in *C. elegans* and found that circRNAs were abundant in the neurons. Interestingly, circRNAs showing higher levels in the neurons tend to be derived from genes that also show higher expression in the neurons. Similar findings were reported in cultured neuron cells, where the increase in circRNA expression is often coupled to upregulation of their linear mRNAs during neuron differentiation (*52*). The time between egg to L1 is the first main period of neuron development in *C. elegans*. At the time of hatching, the majority of neurons (222/302) are already formed (*176*). The high levels of these circRNAs may be due to the active expression of their parental genes for neuron development at the L1 stage. However, there are exceptions. For example, the gene *tbc-17* is a neuronal gene and produces two circRNAs from different exons. One of the two circRNA shows a positive correlation with its mRNA, but the other is depleted in the neurons (Figure 5.3C). Another example *circ-pig-1*, whose cognate mRNA is depleted in sorted samples, is enriched in the neurons (Figure 5.3C).

4.3 RCMs' roles in back-splicing and exon-skipping

Intronic sequences are important for circRNA formation. Especially, Alu repeats in humans are abundant in circRNA-flanking introns (*37*). In *Drosophila*, there is no enrichment of complementary sequences in circRNA-flanking intron pairs (*68*). Rather, long flanking introns are strongly biased for circRNA formation (*68*). In *C. elegans*, circRNA-flanking introns have both features: much more RCM sequences and much longer lengths.

For the first time, I validated that RCMs are required for circRNA formation in multiple circRNA genes. This provides a good method to knock out or knock down circRNAs in *C. elegans*. Particularly, RNAi in *C. elegans* produces secondary siRNAs that recognize sequences other than the primary targets in the same genes (*177*). The only difference in sequences between circRNAs and circRNA-producing exons in their linear mRNAs is the

BSJ sequences. Hence, even though siRNAs specific to the BSJ sequences are used, the secondary siRNAs produced will target the other sequences shared by the circRNAs and their cognate mRNAs, probably making circRNA-specific knockdown (KD) by RNAi not working in *C. elegans*. Except for developing the Cas13-based KD method (*178, 179*), disrupting RCMs sequences may be the only choice to disturb circRNA expression in *C. elegans*. Fortunately, CRISPR-Cas9 based genome editing is quite versatile and highly efficient in *C. elegans* (*180*). In the trials to make RCM deletions in the six circRNA genes, the ratios of picked F1 progeny that showed different genotypes than wild-type were between 33% (5/15, *unc-75*) to 62% (10/16, *gpa-1*) (Figure 5.5), which means that there is a high probability of obtaining edited strains in less than five picked F1 progenies. Sometimes, only one injected P0 worm was needed to obtain the target genome editing, making screening much less labor-intensive.

In *zip-2*, two short pairs of RCMs, 7 nt and 13 nt in length, were identified. To my best knowledge, they are the shortest identified *cis* elements that promote circRNA formation. These RCMs were filtered off in the autoBLAST algorithm (*50*), which was used for global RCM analysis in all circRNA introns (Figure 3.4B and C). This reminds us that special care is needed to identify *cis* elements that regulate circRNA formation when dealing with specific circRNA genes because short RCMs can also be crucial. Moreover, the 13-nt RCMs are highly conserved in the *zip-2* ortholog genes in five nematode species, suggesting their roles in promoting exon-skipping and back-splicing could be conserved.

The level changes of *zip-2* transcripts in the ss/BP strains reflected some interesting results (Figure 3.11C). These results do not affect the conclusion that RCMs directly promote both back-splicing and exon-skipping at the same time. However, further experiments may be worth trying to check whether the increased levels of *L-zip-2* after mutation of ss/BP in intron 1 are due to transcription enhancement or stability enhancement. Many examples in plant or mammalian cells have shown that upstream introns near promoter regions can enhance transcription, which is known as the intron-mediated enhancement (IME) (reviewed in (*181, 182*)). However, here in *zip-2*, disturbance of ss/BP in intron 1 resulted in increases in *L-zip-2* and the skipped or circular transcript, not decrease. Moreover, deletion of RCM sequences in intron 1 did not increase *L-zip-2* (Figure 3.9G), which suggests that the sites involved in splicing (ss/BP) are important, not the other sequences in intron 1.

4.4 A new explanation to the correlation between exon-skipping and back-splicing

Currently, two models have been proposed to explain the correlation between exon-skipping and circRNA formation (*37, 43*): 1). RCM-promoted back-splicing produces circRNAs and y-shaped intermediates, which are further spliced to form skipped transcripts; 2). Exon-skipping produces skipped transcripts and lariat intermediates, which are further back-spliced to form circRNAs (Figure 3.12). The former is used for RCM-driven circRNA genes, and the latter pathway is for circRNA genes that lack RCM sequences. Here, for the first time, I show that RCM sequences in circRNA introns can simultaneously promote both back-splicing and exon-skipping. I further delineated that RCMs are not promoting exon-skipping through back-splicing, neither the other way. Instead, the two pathways are happening together, directly promoted by RCM sequences. I propose that RCMs in the introns not only bring the splice sites for back-splicing to proximity but also bring the sites for exon-skipping together, facilitating both processes simultaneously. Since the RCMs still exist in the intermediates of back-splicing and exon-skipping, they may function twice to promote further splicing/back-splicing in these intermediates (Figure 3.12).

Previous studies of RCMs' roles in circRNA regulation (*31, 48*) or splice sites required for back-splicing (*29*) were mainly based on plasmids in cultured cells. In this thesis, I show that *C. elegans* is a valuable model for *in vivo* investigation of circRNA biogenesis.

4.5 Function of neuronal circRNAs

The initial idea of obtaining the neuronal circRNA profile was to identify potentially functional circRNAs in the neurons. However, the circRNA-KO strains generated in this thesis did not show any obvious or stable phenotypes in several assays related to neuronal functions, like locomotion, chemotaxis, lifespan, and aldicarb resistance. The criterion used for circRNA selection was merely based on expression levels, which may not be reasonable enough. Currently identified functional circRNAs are more based on phenotype-driven screening, or reverse screening, where circRNAs related to a specific phenotype are defined first, which are further screened to identify the functional ones. This approach always requires circRNA profile comparison between assays with minimal differences in genetic background. For example, focusing on differentially expressed and potentially translatable circRNAs between glioblastoma and normal neuron samples, Zhang Nu's lab reported multiple functional short proteins/peptides encoded by circRNAs involved in glioblastoma tumorigenicity (*55, 112-115*). Based on circRNA profile during aging in *C. elegans* (*50*), researchers in Dr. Pedro Miura and Dr. Alexander Van der Linden's lab found a circRNA (*circ-crh-1*) regulating the lifespan (personal communication). Forward screening may be approachable if Cas13-based circRNA KD works well in *C. elegans*, making high-throughput screening possible.

Moreover, even if some phenotypes were observed in some circRNA-KO strains, they should be carefully interpreted because the linear mRNA levels could also be changed (Figure 3.5E). In such cases, rescue of observed phenotypes by re-expression of circRNAs in the KO strains should be necessary.

4.6 *Trans* elements involved in circRNA regulation

Many RBPs are involved in the regulation of splicing. Back-splicing is another choice in alternative splicing (AS). Hence, any factors that regulate other types of AS (exon-skipping, alternative 5'ss or alternative 3'ss, etc.) can be involved in back-splicing regulation. With this logic in mind and using circRNAs in the neurons as targets, I performed a small-scale *in vivo* screening of thirteen conserved RBP genes in their roles of circRNA regulation. Most of these RBPs showed promotional roles in circRNA production, suggesting that the involvement of RBPs in back-splicing may be common in *C. elegans*.

Currently, the only methods to quantify circRNA levels in *C. elegans* are by RT-qPCR or Northern Blot (NB). There are several limitations in these methods: 1).Low throughput. Due to the multiple steps needed in RT-qPCR and NB, it is very difficult for high throughput screening. 2). Inability to check cell-specific circRNA levels. RT-qPCR or NB only quantifies average levels of specific targets in the whole worm. One of the advantages of AS investigation in *C. elegans* is that single-cell resolution can be achieved using specific genetic tools (*173, 174, 183, 184*). Fluorescent back-splicing reporter plasmids have been reported in culture cells using IRES-driven translation of circular reading frame of fluorescent proteins (*34, 59, 60*). Using this fluorescent back-splicing reporter, Li, et al. performed genome-wide RNAi screening of factors involved in circRNA regulation (*59*). This fluorescent back-splicing reporter requires efficient circRNA expression and IRES-

dependent translation, neither of which have been investigated in *C. elegans*. It can be worth trying for a future direction.

4.7 circRNA regulation by both *cis* and *trans* elements

When *cis* elements are disturbed, circRNA levels become undetectable (*circ-glr-2*, *circ-gpa-1*, *circ-unc-75*, *circ-arl-13*, and *circ-iglr-3*) or extremely low (*circ-Y20F4.4* and *circ-zip-2*, < 0.1% remaining) (section 3.2.2). For these circRNAs, their levels can be either upregulated or downregulated by RBPs, which are *trans* elements. The level changes caused by loss of RBPs were quite limited, ranging from 0.2- to 2.2-fold changes (Figure 3.13A). These results suggest a theory for circRNA formation, where *cis* elements such as RCM sequences determine whether one circRNA can be produced (yes or no), and *trans* elements, mainly RBPs, fine-tune the levels of circRNAs, rendering the cell-specific expression of circRNAs possible (less or more).

4.8 Mechanism of FUST-1 in back-splicing and exon-skipping regulation

FUST-1 can regulate both exon-skipping and back-splicing in the same genes (Figure 3.16). In *zip-2*, the 5'ss of exon-skipping and back-splicing are important for FUST-1's role in back-splicing and exon-skipping, respectively (Figure 3.17). U1 snRNA could be involved in the interaction between FUST-1 and the spliceosome complex (Figure 3.17). It is helpful to understand the mechanisms of FUST-1 regulation if the binding sites in each gene can be mapped. CLIP-seq data on FUS suggest that rather than recognizing specific sequences, FUS tends to bind to stem-loop secondary structures (*164, 165, 185, 186*). In cultured mouse neuroblastoma N2a cells, FUS binds to the flanking introns of circularized exons (*61*). For FUST-1, although it is homologous to FUS, its binding sites in *C. elegans* may show different patterns. Mapping the binding sites of FUST-1 *in vivo* using CLIP-seq technologies, iCLIP (*187*) or eCLIP (*188, 189*), can be used for further understanding of the mechanisms of FUST-1 in regulating exon-skipping and back-splicing.

4.9 FUST-1 knock-in models

To further investigate the relationship between natural mutations in FUS and circRNA regulation, two FUST-1 mutations (R446S and P447L) were used to mimic two types of mutations in the NLS region (R524S and P525L) of FUS. Of these two FUST-1 mutations, only P447L dramatically affected nuclear localization. However, circRNAs were significantly affected in both strains, indicating some other functions by the NLS in the C-terminal of FUST-1.

Consistent with the findings that P525L-mutated FUS could not rescue altered circRNAs in FUS-knockdown N2a cells (*61*), P447L-mutated FUST-1 showed malfunction in circRNA regulation, suggesting that the mutation of conserved sites in FUST-1 can somehow reflect the role of the mutations in FUS. Hence, other than knock-in of human FUS (wild-type or mutated isoforms) in *C. elegans* (*190-195*), the conserved residues in FUST-1 can be used as targets to generate *C. elegans* models to investigate FUS mutations (*172*).

4.10 The autoregulation loop in *fust-1*

Self-regulation has been reported in FUS, where FUS promotes skipping of exon 7 of its pre-mRNA, which results in NMD (*186*). Unlike the previous example, FUST-1a-promoted exon skipping of *fust-1* pre-mRNA produces FUST-1b that contains exactly the same functional domains, but with different N-terminal sequences. While FUST-1a is capable of promoting exon-skipping and circRNA regulation, FUST-1b is not functional in either of the two aspects (Figure 3.19F and Figure 3.20A). I first hypothesized that N-terminal sequences in FUST-1a might be important for its function. However, the FUST-1- ΔN construct, which has no N-terminal sequences, seemed functional in both exon-skipping promotion and circRNA regulation, although not as efficient as FUST-1a. These results suggest that N-terminal in FUST-1b may interfere with the functional domain(s), possibly RRM, so that FUST-1b cannot bind to the target mRNAs recognized by FUST-1a.

The frameshifting in FUST-1b dramatically changes the N-terminal amino acid contents compared with FUST-1a. The isoform a-specific N-terminal has high ratios of glycine (53/164, 32.3%) and glutamine (22/164, 13.4%). However, isoform b-specific N-terminal contains more valine (18/106, 17.0%) and glutamic acid residues (18/106, 17.0%), which are very few in isoform a-specific N-terminal: 0/164 and 2/164, respectively. High valine content may make the N-terminal of FUST-1b more hydrophobic and high glutamic acid content can add more negative charges, which may cause the folding of FUST-1b different from FUST-1a. Further *in vitro* RNA binding experiments or structural analysis may be worth trying to investigate the detailed mechanisms that dictate different function potentials in the two FUST-1 isoforms.

4.11 Summary

In this thesis, I investigated the regulation of circRNAs by both *cis* elements and *trans* elements in *C. elegans*. Starting from large-scale neuron sorting from L1 worms, I obtained the first neuronal circRNA profile in *C. elegans*, where circRNA levels showed a strong positive correlation to their cognate mRNA levels. For *cis* elements, using identified circRNAs with high expression levels in the neurons, I validated that RCM sequences are required for circRNA formation. Moreover, RCMs also promote correlated skipping events in multiple circRNA genes. Using *zip-2* as a model gene with explicit expressions of both circular and skipped transcripts, I showed that RCMs directly promote both exon-skipping and back-splicing simultaneously, which provides a new explanation to the correlation between exon-skipping and circRNA formation.

Regarding *trans* elements involved in circRNA regulation, I identified FUST-1 as a circRNA regulator by screening thirteen conserved RBPs. When recognizing the pre-mRNAs of circRNA genes, FUST-1 can regulate the correlated exon-skipping and back-splicing at the same time. In *zip-2*, the 5' splice sites for back splicing and exon skipping are important for FUST-1 to regulate exon-skipping and back-splicing, respectively. I also showed that circRNAs were dysregulated in two FUST-1 knock-in strains with a mutation in the NLS region (R446S and P447L).

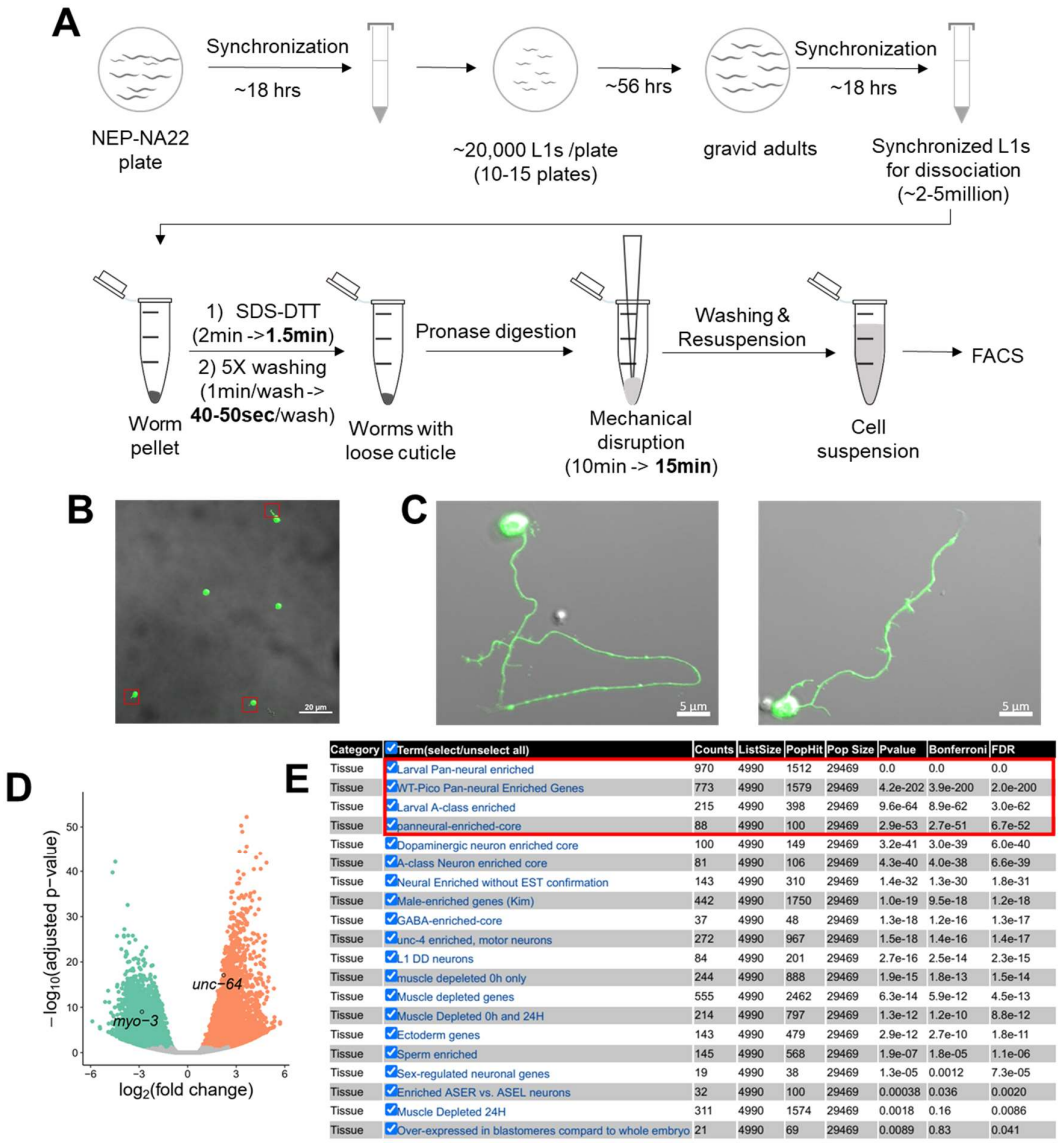
Finally, I identified a new autoregulation loop in *fust-1* that is important for its role in circRNA regulation. FUST-1a promotes the skipping of exon 5 of its own pre-mRNA, which produces FUST-1b. I confirmed that FUST-1a is functional in circRNA regulation. Although with the same functional domains (RRM, ZnF, and NLS), FUST-1b cannot promote exon-skipping of *fust-1* pre-mRNA; neither can it promote circRNA formation.

This thesis provides new perspectives in the understanding of circRNA regulation *in vivo*, where RCM sequences (*cis* elements) determine whether circRNA can be formed or not, and

RBPs (*trans* elements) regulate how much they can be produced. Further, the circRNA regulation network can be integrated with pathways involved in the production of functional RBP isoforms.

I hope the results in this thesis can attract people in the circRNA field to use *C. elegans* as an *in vivo* model for circRNA investigation, and promote people in the *C. elegans* field to develop tools for circRNA study, like Cas13-based system for circRNA-specific KD, fluorescent back-splicing reporter system, etc.

5. Supplementary data



Category	Term(select/unselect all)	Counts	ListSize	PopHit	Pop Size	Pvalue	Bonferroni	FDR
Tissue	Larval Pan-neural enriched	970	4990	1512	29469	0.0	0.0	0.0
Tissue	WT-Pico Pan-neural Enriched Genes	773	4990	1579	29469	4.2e-202	3.9e-200	2.0e-200
Tissue	Larval A-class enriched	215	4990	398	29469	9.6e-64	8.9e-62	3.0e-62
Tissue	panneural-enriched-core	88	4990	100	29469	2.9e-53	2.7e-51	6.7e-52
Tissue	Dopaminergic neuron enriched core	100	4990	149	29469	3.2e-41	3.0e-39	6.0e-40
Tissue	A-class Neuron enriched core	81	4990	106	29469	4.3e-40	4.0e-38	6.6e-39
Tissue	Neural Enriched without EST confirmation	143	4990	310	29469	1.4e-32	1.3e-30	1.8e-31
Tissue	Male-enriched genes (Kim)	442	4990	1750	29469	1.0e-19	9.5e-18	1.2e-18
Tissue	GABA-enriched-core	37	4990	48	29469	1.3e-18	1.2e-16	1.3e-17
Tissue	unc-4 enriched, motor neurons	272	4990	967	29469	1.5e-18	1.4e-16	1.4e-17
Tissue	L1 DD neurons	84	4990	201	29469	2.7e-16	2.5e-14	2.3e-15
Tissue	muscle depeleted 0h only	244	4990	888	29469	1.9e-15	1.8e-13	1.5e-14
Tissue	Muscle depleted genes	555	4990	2462	29469	6.3e-14	5.9e-12	4.5e-13
Tissue	Muscle Depleted 0h and 24H	214	4990	797	29469	1.3e-12	1.2e-10	8.8e-12
Tissue	Ectoderm genes	143	4990	479	29469	2.9e-12	2.7e-10	1.8e-11
Tissue	Sperm enriched	145	4990	568	29469	1.9e-07	1.8e-05	1.1e-06
Tissue	Sex-regulated neuronal genes	19	4990	38	29469	1.3e-05	0.0012	7.3e-05
Tissue	Enriched ASER vs. ASEL neurons	32	4990	100	29469	0.00038	0.036	0.0020
Tissue	Muscle Depleted 24H	311	4990	1574	29469	0.0018	0.16	0.0086
Tissue	Over-expressed in blastomeres compard to whole embryo	21	4990	69	29469	0.0089	0.83	0.041

Figure 5.1 Worm dissociation and neuron sorting. Modified from (*148*). (A) Steps of L1 worm preparation and dissociation for FACS. Optimized conditions are in bold. (B) Representative confocal image of sorted cells. Note the short neurites (red rectangles) of some cells. Scale bar: 20 µm. (C) Confocal images showing sorted neurons after five-day culture at 20°C. Scale bars are 5 µm. (D) Volcano plot showing differentially expressed genes between the sort group and the whole group. *myo-3* and *unc-64* are labeled. (D) Output from WormExp for gene set enrichment search using upregulated genes in the neurons in our dataset. The red rectangle highlights the top 4 hits.

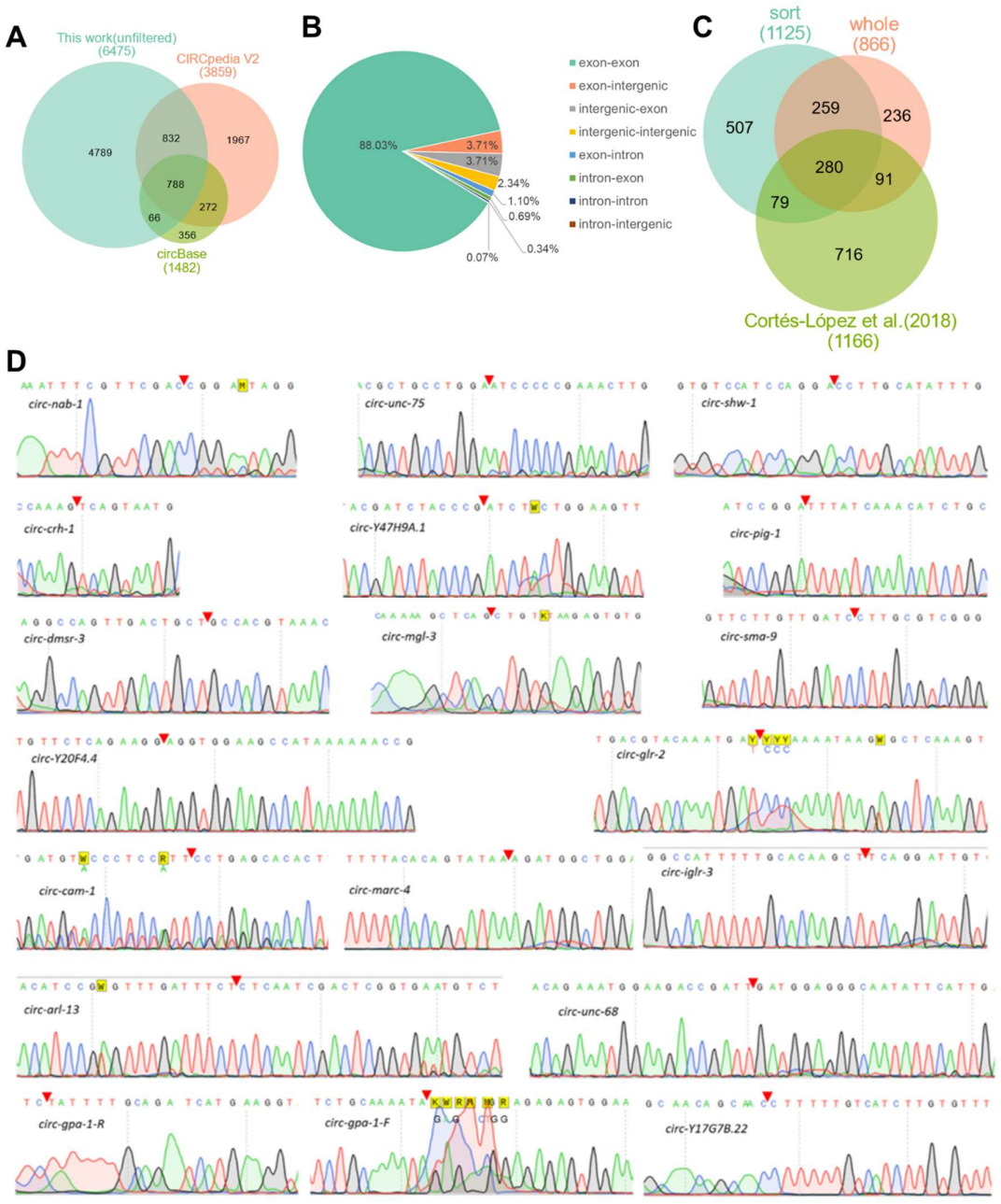


Figure 5.2 circRNA annotation and validation. Modified from (*148*). (A) Overlap of circRNAs identified in this work and circRNAs of *C. elegans* in two databases (CIRCpedia V2 and circBase). (B) Ratios of junction types of filtered circRNAs. (C) Overlap of filtered circRNAs in this work and filtered circRNAs in work of Cortés-López et al (*50*). (D) Sanger sequencing results of the BSJ sequences of selected circRNAs. Red triangles denote the joint sites.

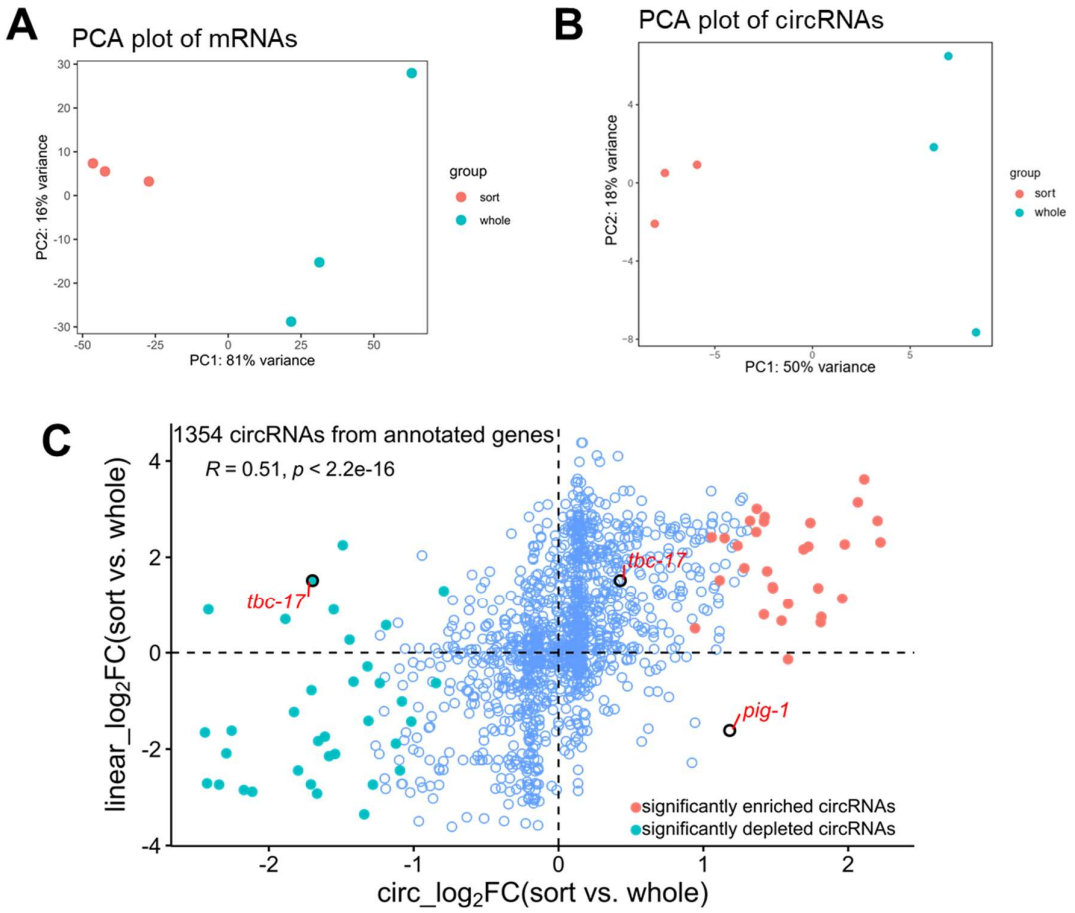


Figure 5.3 circRNA expression analysis. (A, B) PCA plot of mRNAs (A) and circRNAs (B) in the sort group and the whole group. (C) Scatter plot showing the correlation of log2 fold change of circRNAs and their cognate linear RNAs in the sort group and whole group. Gene names of some circRNAs were labeled. The Pearson correlation coefficient (*R*) and *p* value (*p*) are shown. Significantly DE circRNAs are shown by colored dots.

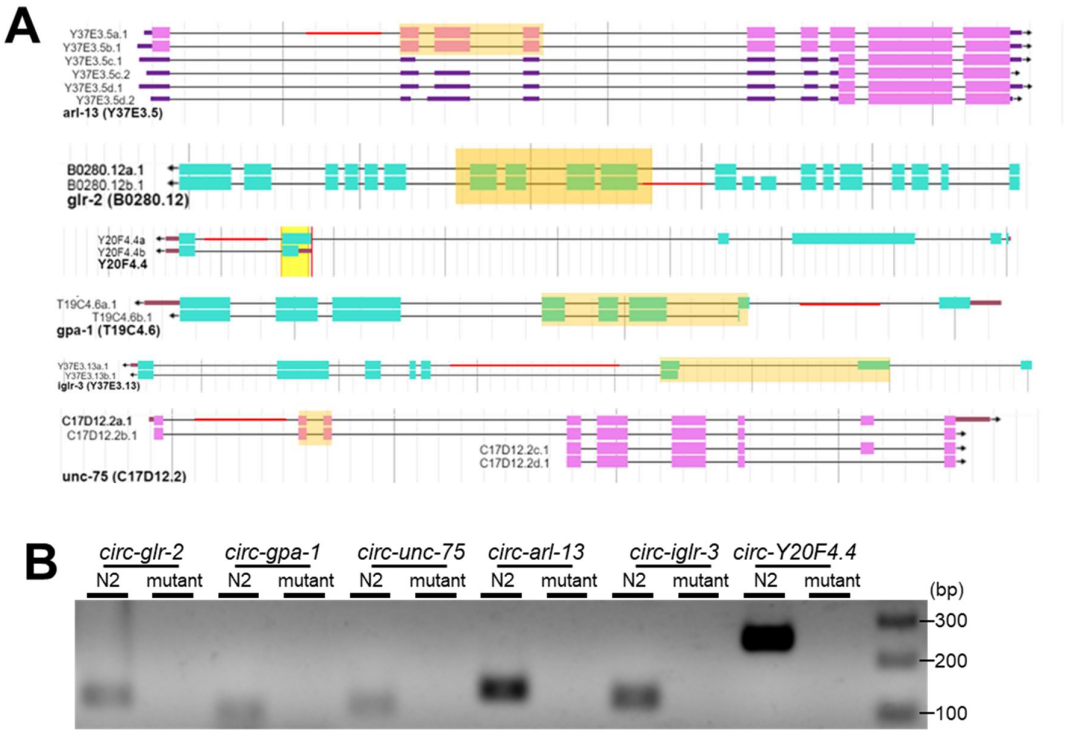


Figure 5.4 RCM deletion by CRISPR-Cas9. Modified from (*148*). (A) Position of deleted RCMs in six circRNA genes. (B) RT-PCR detection of circRNAs in wild-type N2 strain and the RCM-deleted (mutant) strains. Note that all circRNAs cannot be amplified in the mutant strains.

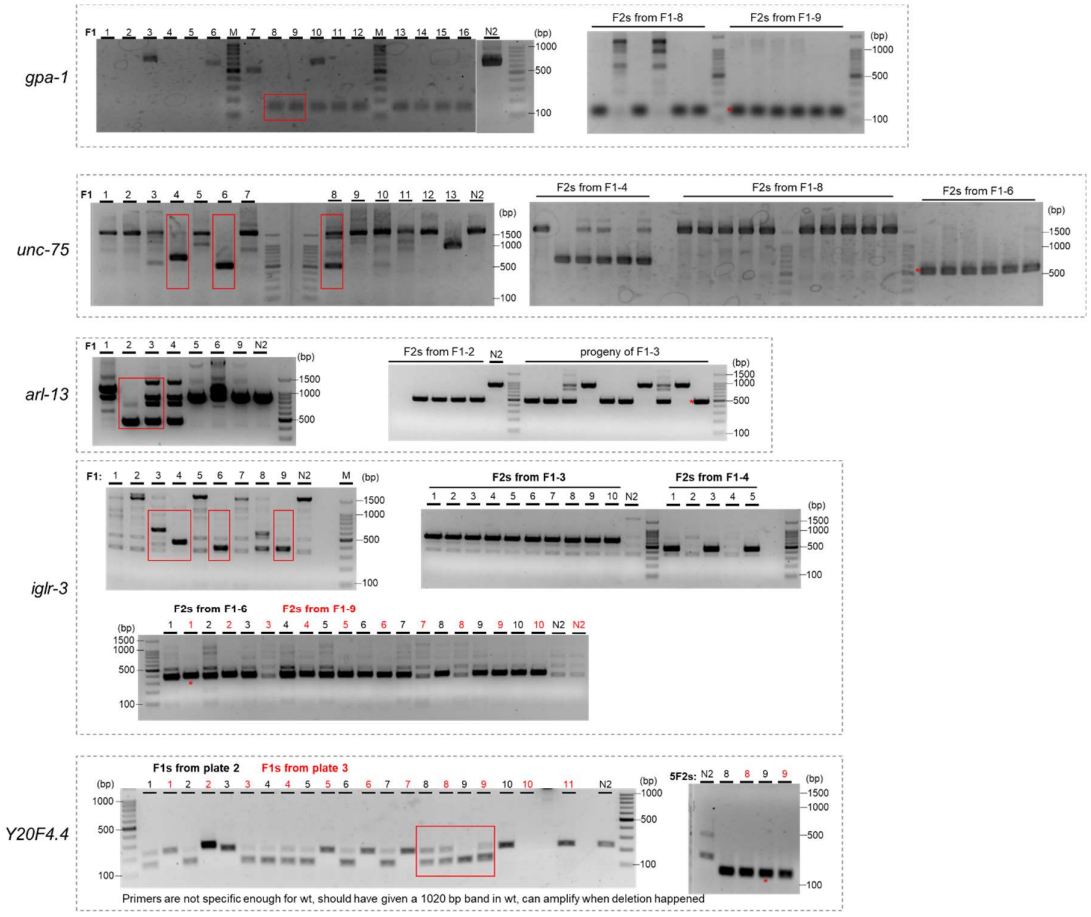


Figure 5.5 Genotype screening of F1 and F2 worms in the deletion of RCMs in the other five circRNA genes. Red rectangles show the strains used for F2 screening. Asterisks indicate strains with the target deletions, which were kept as circRNA knockout strains.

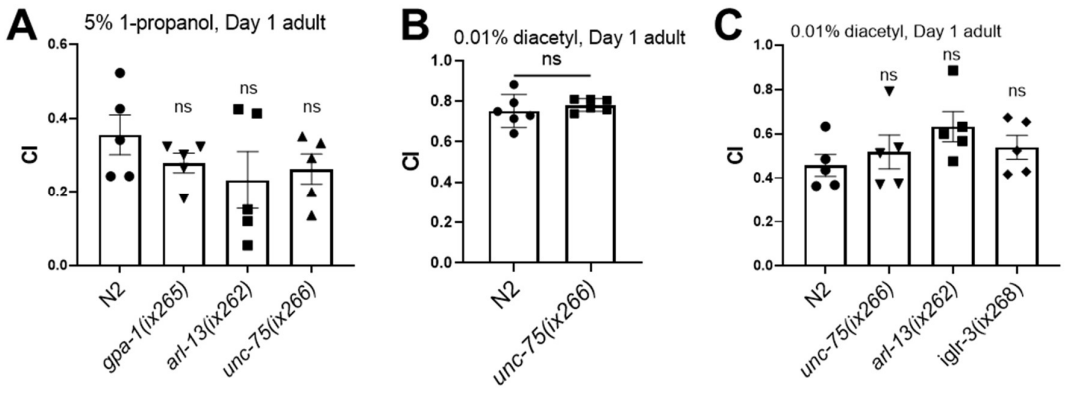


Figure 5.6 Comparisons of chemotaxis index towards 5% propanol or 0.01% diacetyl between wild-type N2 strain and circRNA KO strains. A, B and C show independent trails on different days. (A, C) ns, not significant, one-way ANOVA with Dunnett's multiple comparisons with N2 group. (B) ns, not significant, two-tailed Student's *t*-test.

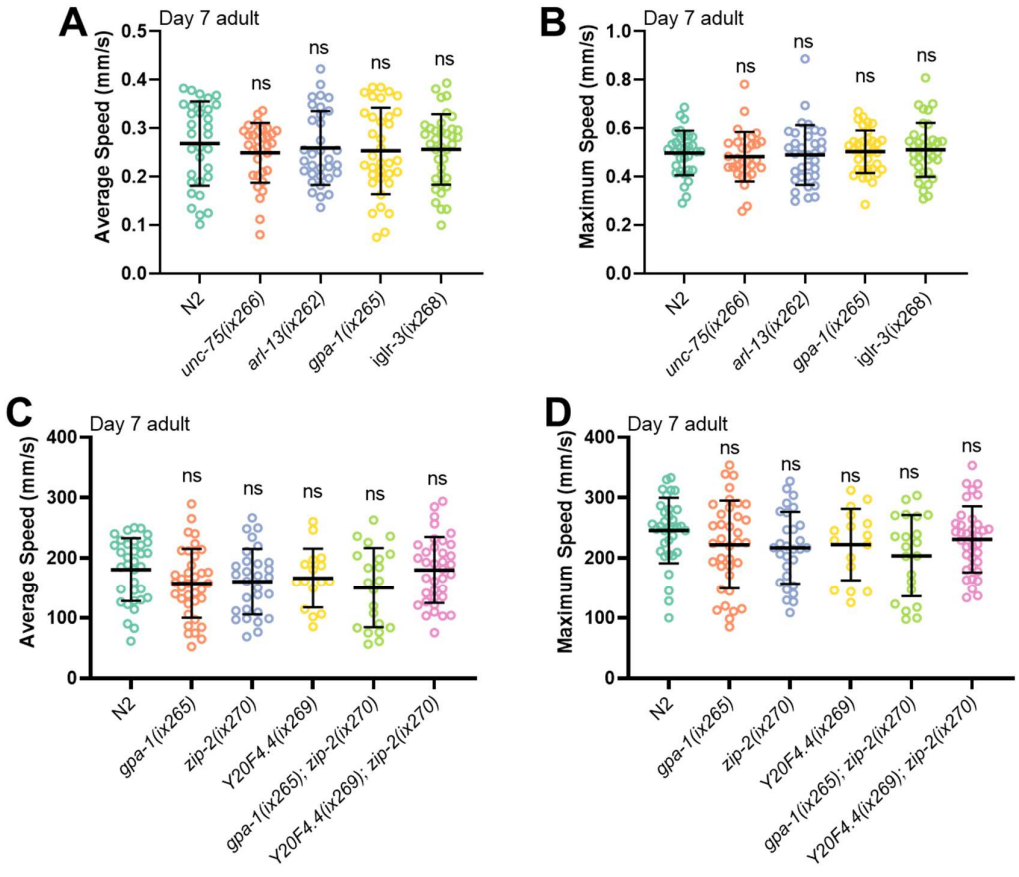


Figure 5.7 Locomotion speed comparisons between wild-type N2 strain and circRNA KO strains. (A, C) Average speed of indicated strains at day-7 adult stage. (B, D) Maximus speed of indicated strains at day-7 adult stage. (A, B, C, D) ns, not significant, one-way ANOVA with Dunnett's multiple comparisons with the N2 group.

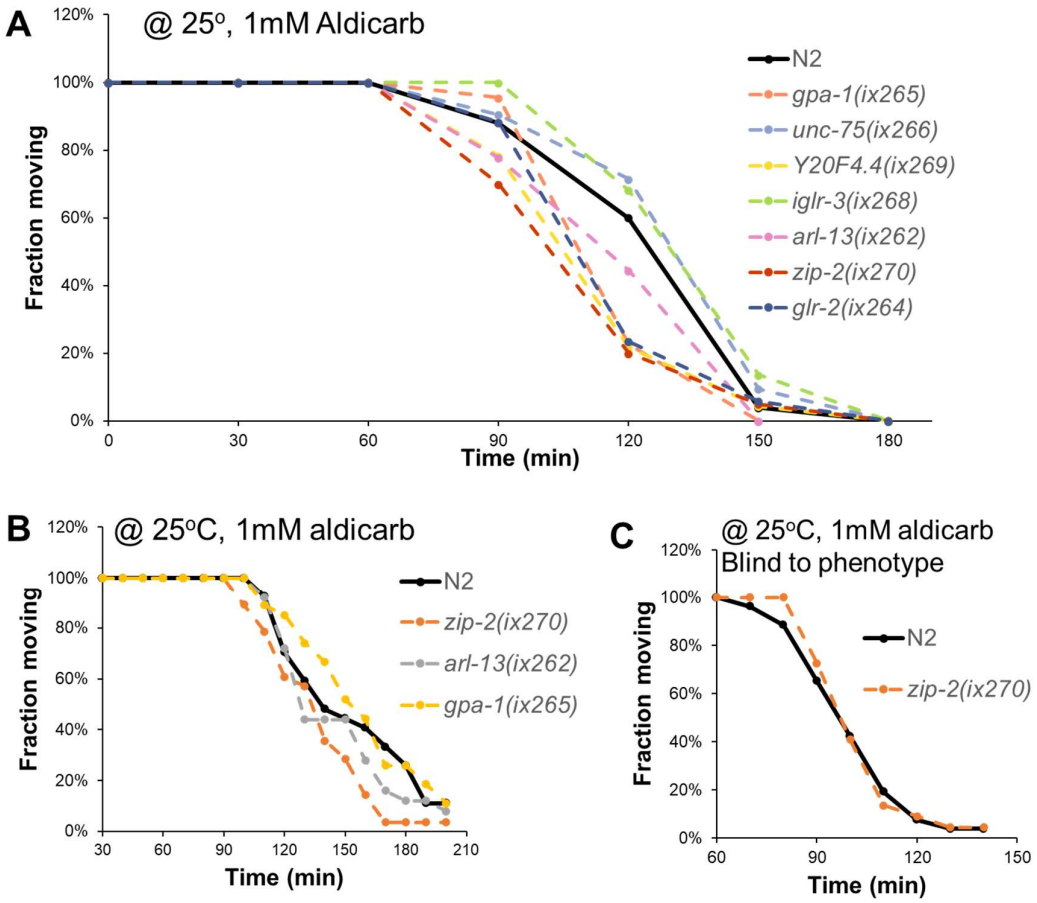


Figure 5.8 Results of aldicarb resistance assays. (A, B) Independent trials without being blind to phenotypes. *zip-2(ix270)* seemed to be less resistant to 1 mM aldicarb. (C) A trail being blind to the genotypes of tested strains. No difference between the two strains was observed.

Table 5.1 Raw data of aldicarb resistance assay shown in Figure 5.8A.

Time (min)	N2	*gpa-1 (ix265)*	*unc-75 (ix266)*	*Y20F4.4 (ix269)*	*iglr-3 (ix268)*	*arl-13 (ix262)*	*zip-2 (ix270)*	*glr-2 (ix264)*
0	25	22	21	23	22	18	20	17
30	25	22	21	23	22	18	20	17
60	25	22	21	23	22	18	20	17
90	22	21	19	18	22	14	14	15
120	15	5	15	5	15	8	4	4
150	1	0	2	1	3	0	1	1
180	0		0	0	0		0	0

Table 5.2 Raw data of aldicarb resistance assay shown in Figure 5.8B.

Time (min)	N2	*zip-2(ix270)*	*arl-13(ix262)*	*gpa-1(ix265)*
30	27	28	25	27
40	27	28	25	27
50	27	28	25	27
60	27	28	25	27
70	27	28	25	27
80	27	28	25	27
90	27	28	25	27
100	27	25	25	27
110	25	22	23	24
120	19	17	18	23
130	16	16	11	20
140	13	10	11	18
150	12	8	11	14
160	11	4	7	12
170	9	1	4	7
180	7	1	3	7
190	3	1	3	5
200	3	1	2	3

Table 5.3 Raw data of aldicarb resistance assay shown in Figure 5.8C.

Time (min)	N2	*zip-2(ix270)*
60	26	22
70	25	22
80	23	22
90	17	16
100	11	9
110	5	3
120	2	2
130	1	1
140	1	1

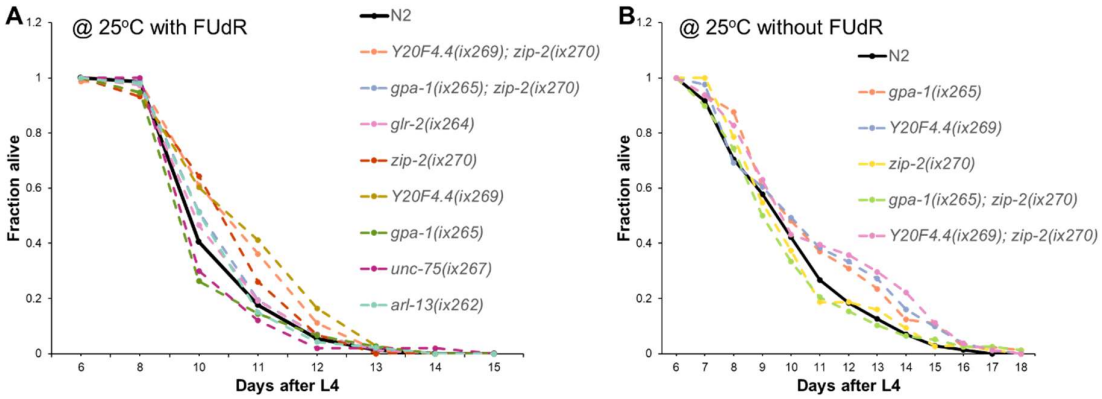


Figure 5.9 Results of lifespan assays. (A) Lifespan comparisons in the indicated strains. 50 μM FUdR (5'-fluorodeoxyuridine) was used. (B) Lifespan comparisons without FUdR.

Table 5.4 Raw data of lifespan assay shown in Figure 5.9A.

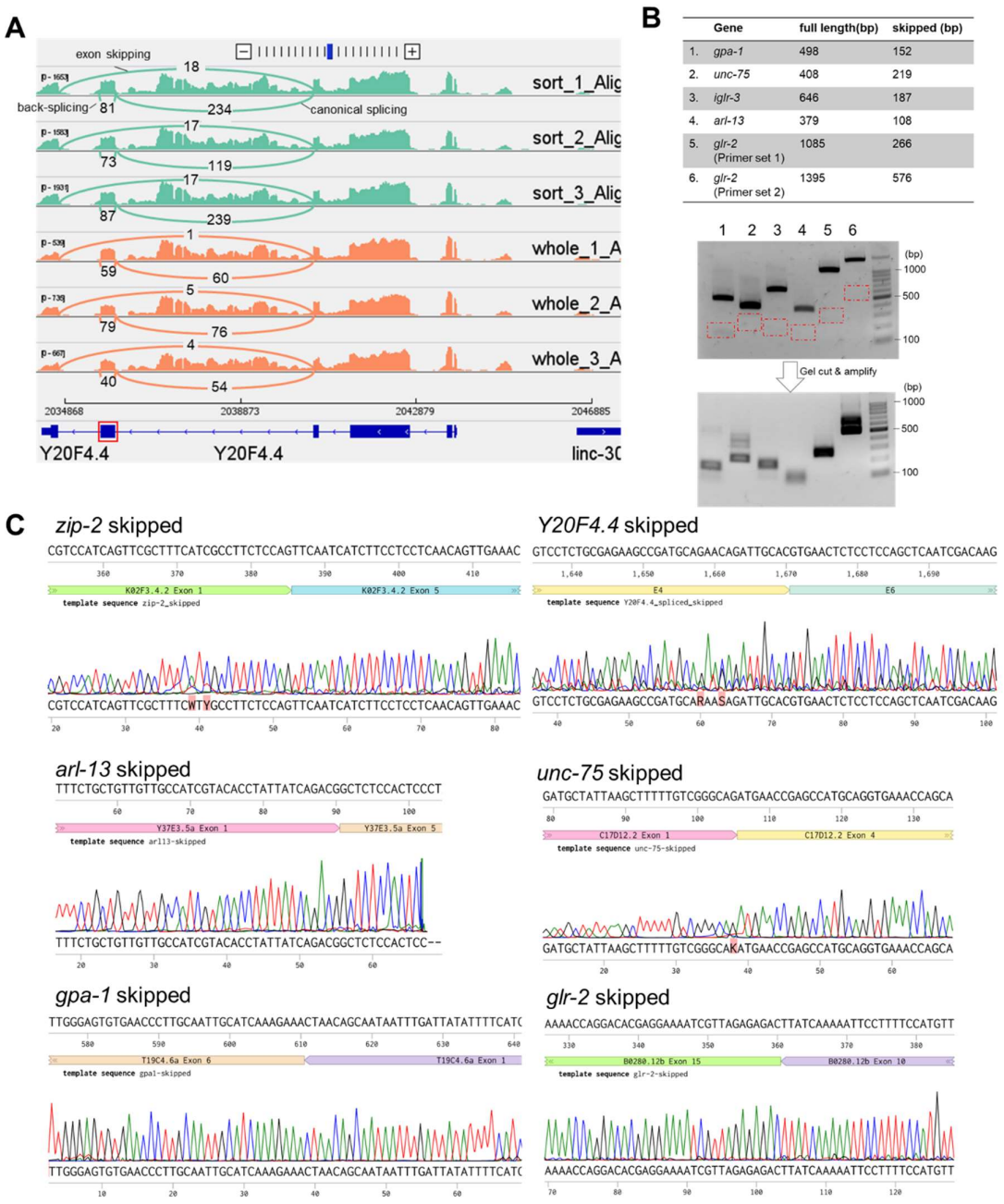
Strain	Rep.	Days after L4								Sum
		6	8	10	11	12	13	14	15	
N2	1	0	0	16	6	2				24
	2	0	0	13	6	4	1	1		25
	3	0	1	14	5	3	2			25
Y20F4.4(ix269); zip-2(ix270)	1	1	0	7	7	8	2			25
	2	0	0	9	6	5	3			23
	3	0	0	11	5	5	2	1		24
gpa-1(ix265); zip-2(ix270)	1	0	0	15	5	3	0	1		24
	2	0	1	10	9	2	1			23
	3	0	0	9	9	4	3			25
glr-2(ix264)	1	0	0	11	7	3	3			24
	2	0	1	13	8	2	1			25
	3	0	1	13	5	4	0	1		24
zip-2(ix270)	1	0	1	8	8	4	4			25
	2	0	1	7	11	5	1			25
	3	0	3	6	9	5				23
iglr-3(ix268)	1	0	2	12	7	3	1	0	1	26
	2	2	1	15	4	0	2	1		25
	3	1	3	11	6	1	0	2		24
Y20F4.4(ix269)	1	0	0	7	4	9	3	1		24
	2	0	1	11	5	7	1			25
	3	0	3	7	5	2	6	1		24
gpa-1(ix265)	1	0	2	17	2	1	2	1		25
	2	0	1	15	5	3	1	1		26
	3	0	1	20	2	2				25
unc-75(ix267)	1	0	0	18	6	1				25
	2	0	0	17	3	4	0	0	1	25
arl-13(ix262)	1	0	0	9	8	4	1	1		23
	2	0	1	13	9	1				24

Table 5.5 Raw data of lifespan assay shown in Figure 5.9B.

Strain	Rep.	Dead/ Censored	Days after L4												
			7	8	9	10	11	12	13	14	15	16	17	18	Sum
N2	1	D	0	3	1	6	6	2	3	4					25
		C	0	0	0	0	0	0	1	0					1
	2	D	2	8	3	3	4	2	0	0	1	1			24
		C	2	1	0	0	0	1	0	0	0	0			4
	3	D	4	4	5	2	1	2	1	0	2	0	1		22
		C	0	0	1	0	0	0	0	0	0	0	0		1
gpa-1(ix265); zip-2(ix270)	1	D	2	5	7	4	3	1	3	1					26
		C	1	0	0	1	0	0	0	0					2
	2	D	3	5	6	2	2	3	1	0	1	2			25
		C	2	2	0	0	0	0	0	0	0	0			4
	3	D	3	2	6	7	5	0	0	2	0	0	0	1	26
		C	0	0	0	0	0	0	0	0	0	0	0	0	0
Y20F4.4(ix269); zip-2(ix270)	1	D	3	2	6	4	2	1	3	1	5	0	1		28
		C	2	1	0	0	0	0	0	0	0	0	0		3
	2	D	2	1	3	4	1	1	1	4	2	5	1	1	26
		C	0	1	0	2	0	0	0	0	0	0	0	0	3
	3	D	0	6	7	8	0	1	1	1	2	1			27
		C	0	0	1	1	0	0	0	0	1	0			3
zip-2(ix270)	1	D	0	5	4	2	5	0	2	1	0	0	0	1	20
		C	1	0	0	1	0	0	0	0	0	0	0	0	2
	2	D	0	3	7	10	4	0	0	3	2	0	1		30
		C	0	0	0	0	0	0	0	0	0	0	0		0
	3	D	0	8	7	1	5	0	0	1	3	0	0		25
		C	1	1	0	0	0	0	0	0	0	0	1		3
gpa-1(ix265)	1	D	2	2	8	4	4	1	2	3	1	0	0	0	27
		C	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	D	1	2	8	3	2	1	2	3	0	4	0	1	27
		C	1	0	0	0	0	0	0	0	0	0	0	0	1
	3	D	2	1	6	3	3	3	2	3	0	3			26
		C	0	0	0	0	0	0	0	0	0	0			0
Y20F4.4(ix269)	1	D	1	6	3	3	4	0	0	5	3	1	1		27
		C	0	1	0	0	0	1	0	0	0	0	0		2
	2	D	1	8	3	4	2	2	3	2	0	2	0	1	28
		C	2	1	1	0	0	0	0	0	0	0	0	0	4
	3	D	0	9	1	2	3	2	2	2	2	2	1		26
		C	0	0	0	0	0	0	0	0	0	0	0		0

Table 5.6 Log-Rank test of results in Figure 5.9B, produced from Oasis web server (https://sbi.postech.ac.kr/oasis/).

Condition	Statistics		
	Chi^2	P-value	Bonferroni P-value
N2 *vs. gpa-1(ix265); zip-2(ix270)*	0.88	0.3493	1
N2 *vs. Y20F4.4(ix269); zip-2(ix270)*	5.83	0.0158	0.079
N2 *vs. zip-2(ix270)*	0.12	0.7264	1
N2 *vs. gpa-1(ix265)*	2.07	0.1507	0.7535
N2 *vs. Y20F4.4(ix269)*	3.27	0.0707	0.3533



	Gene	full length(bp)	skipped (bp)
1.	*gpa-1*	498	152
2.	*unc-75*	408	219
3.	*iglr-3*	646	187
4.	*arl-13*	379	108
5.	*glr-2* (Primer set 1)	1085	266
6.	*glr-2* (Primer set 2)	1395	576

Figure 5.10 Skipped transcripts in circRNA genes. Modified from (*148*). (A) Sashimi plot showing numbers of reads aligned to the skipped junction (exon-skipping), the back-spliced junction (back-splicing), and the canonical splicing junction between exon 4 and exon 5 (canonical splicing) in *Y20F4.4*.The exon in the red rectangle is to form *circ-Y20F4.4*. (B) Amplification of the skipped transcripts from several circRNA genes by two-round PCRs. Red rectangles mark the gel areas to be cut. (C) Confirmation of sequences of the skipped transcripts in six circRNA genes.

Table 5.7 Enzyme digestion patterns of wild-type and RCM-deleted sequences for genotype screening in *zip-2*.

Target	Enzyme	wt	mutation
intron1-RCM-del	DraI	173 + 114 bp	256 bp (-31 bp)
intron4-RCM-del	DraI	80 + 162 bp	227 bp (-15 bp)

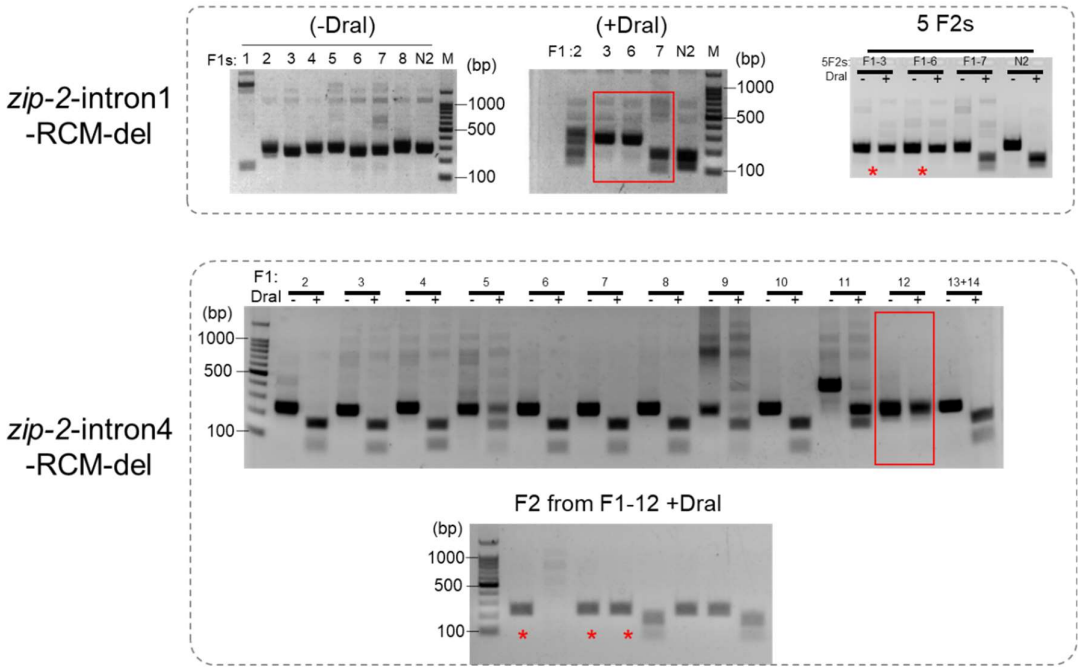


Figure 5.11 Genotype screening of F1 and F2 worms in the deletion of RCMs in *zip-2*. Red rectangles show the F1 strains used for F2 screening. Asterisks indicate strains with the target deletions.

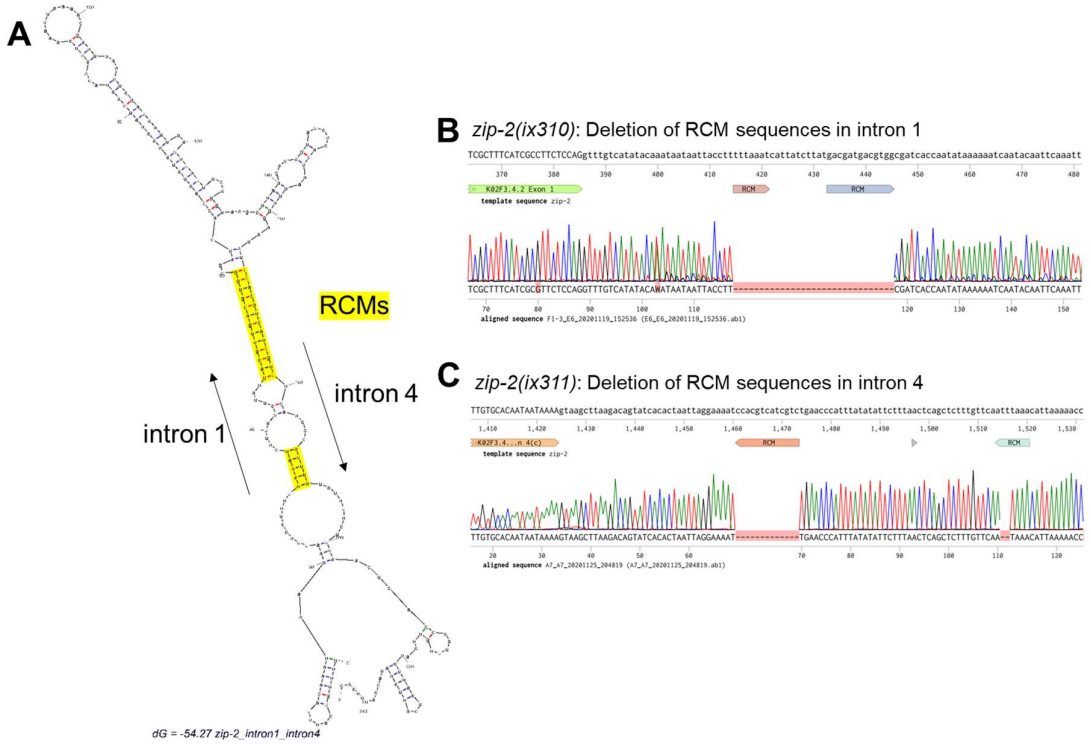


Figure 5.12 RCMs in circRNA-flanking introns of *zip-2*. Modified from (*148*). (A) Folding prediction of intron 1 and intron 4 of *zip-2* by Mfold (http://www.unafold.org/mfold/applications/rna-folding-form.php). RCM sequences are highlighted. (B, C) Deleted RCM sequences in intron 1 and intron 4 of *zip-2*.

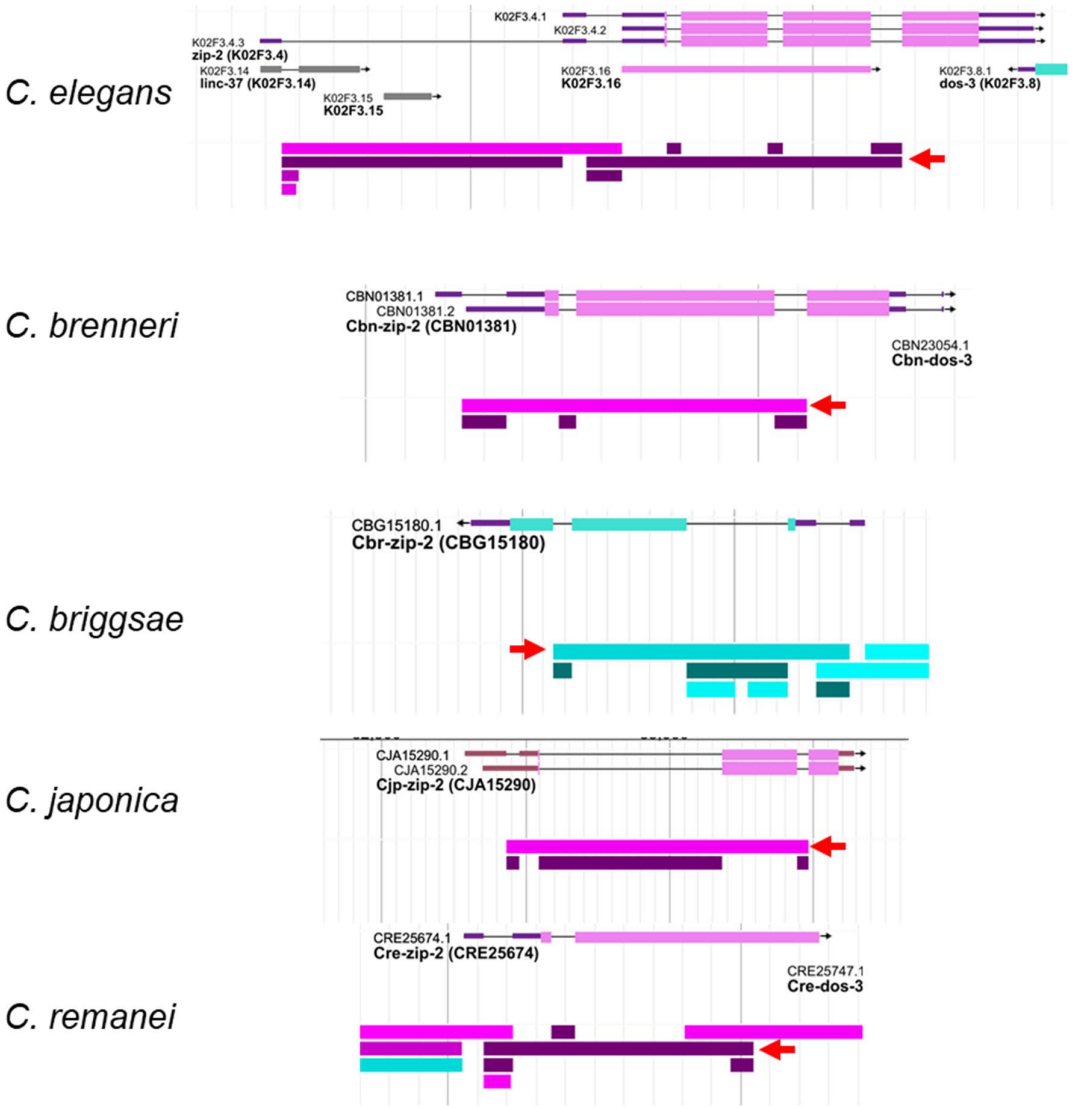


Figure 5.13 Gene structures of *zip-2* ortholog genes. Modified from (*148*). Gene structures of ortholog *zip-2* genes in indicated nematode species are shown. The splicing patterns of these genes are also shown (from WormBase: https://wormbase.org/). Red arrows indicate the splice junctions of the skipped transcripts.

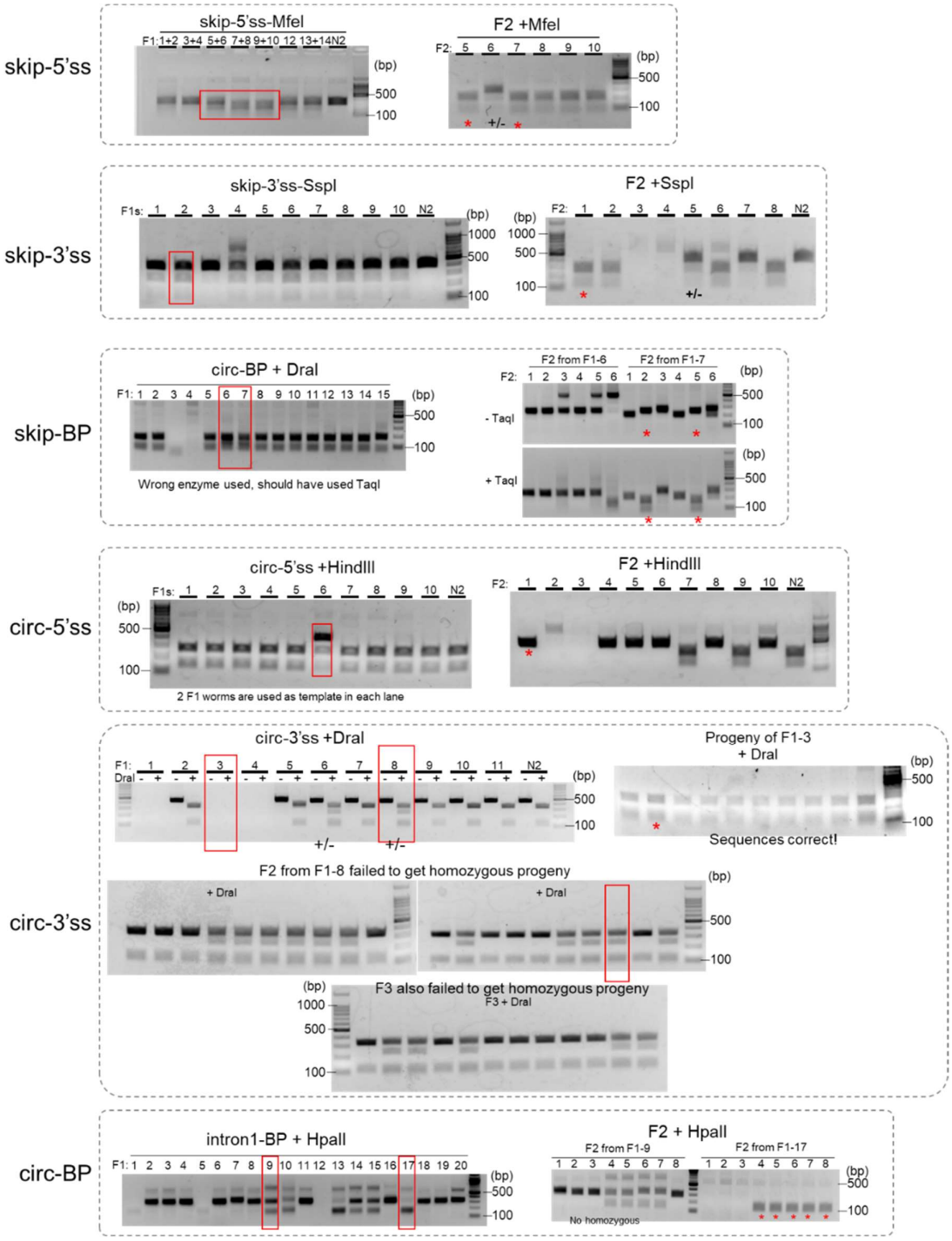


Figure 5.14 Genotype screening of F1 and F2 worms in ss/BP mutations in *zip-2*. Red rectangles show the F1 strains used for F2 screening. Asterisks indicate strains with the designed mutations.

Table 5.8 Enzymes and the expected digestion patterns of wild-type and ss/BP-mutated sequences.

Mutation site	Enzyme	wt	Mutation
skip-5'ss	MfeI	287 bp	82 + 203 bp
skip-3'ss	SspI	400 bp	264 + 133 bp
skip-BP	TaqI	242 bp	158 + 84 bp
circ-5'ss	HindIII	265 +136 bp	400 bp
circ-3'ss	DraI	341 + 114 bp	114 + 91 +248 bp
circ-BP	HpaII	287 bp	151 + 136 bp

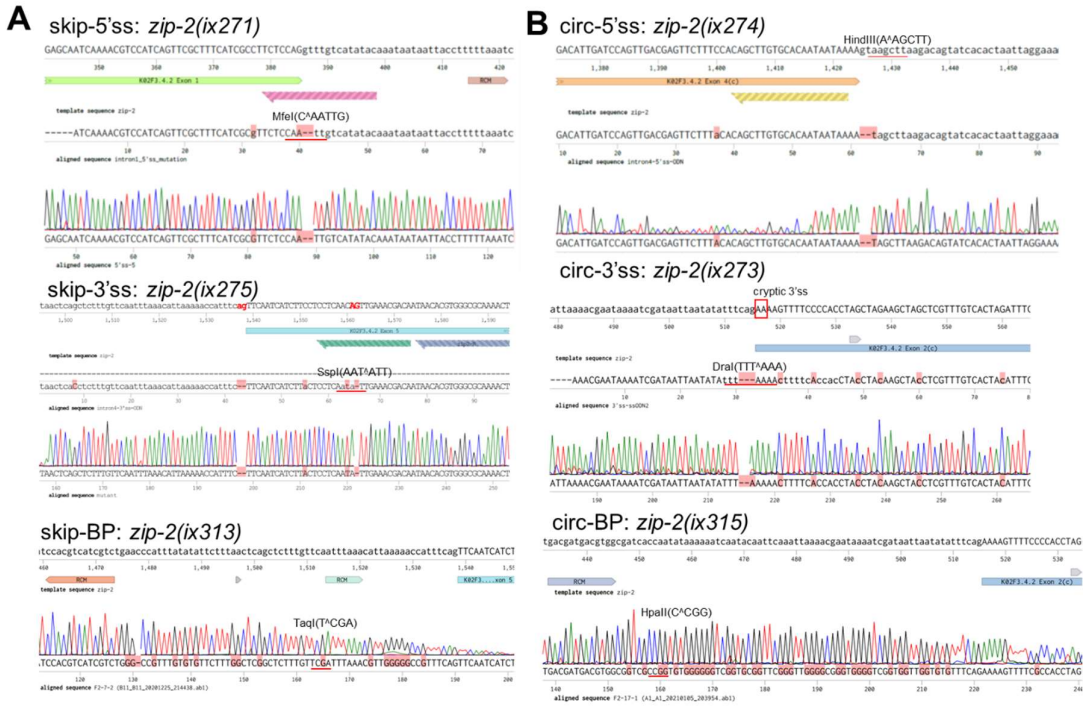


Figure 5.15 Sequence confirmation of mutated ss and BP sites in *zip-2*. Modified from (*148*). (A, B) Sanger sequence results of splicing sites and branch points mutation in intron 1 and intron 4 of *zip-2*. The enzyme digestion sites used to distinguish wild-type sequences and mutated sequences are labeled. The position of the cryptic 3'ss in circ-3'ss strain is labeled.

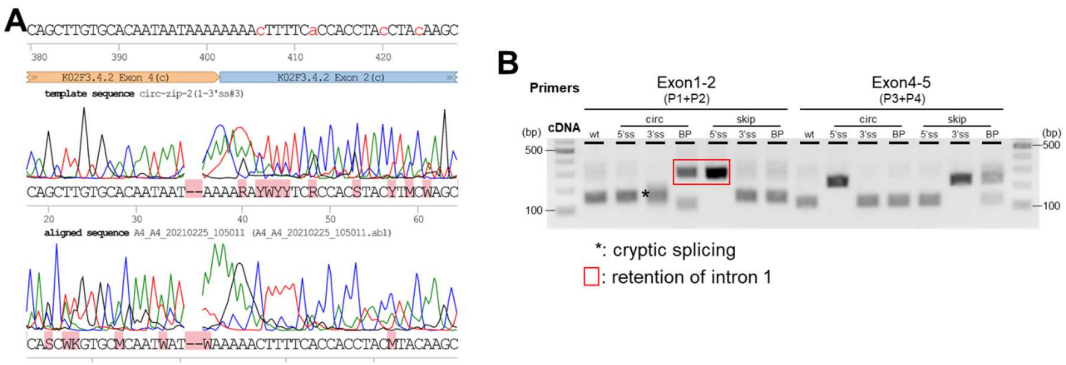


Figure 5.16 Detection of *zip-2* transcripts in ss/BP mutated strains. Modified from (*148*). (A) Sanger sequence of *circ-zip-2* produced from the *zip-2*(circ-3'ss) strain. Note that amplified sequences are 2 nt shorter than the predicted BSJ sequences. (B) Splicing patterns between exon 1 and 2 and between exon 4 and 5 in zip-2 in indicated strains. The asterisk indicates the use of the cryptic splice site in circ-BP mutated strain. The red rectangle highlights the retention of intron 1 in strains with mutation of circ-BP or skip-5'ss.

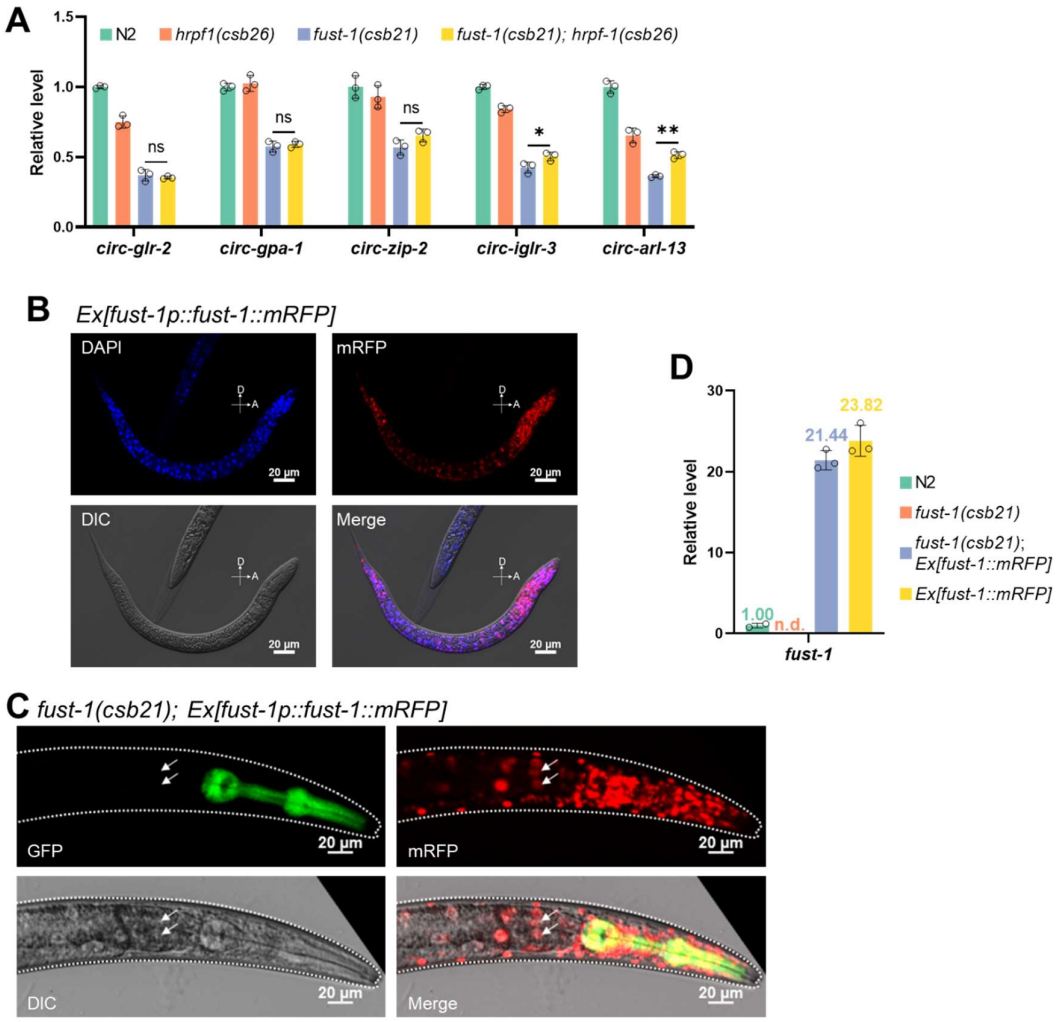


Figure 5.17 Identification of FUST-1 as a circRNA regulator. Modified from (*161*). (A) RT-qPCR quantification of circRNA levels at the L1 stage of indicated strains. Note that there is no additive effect in *fust-1(csb21); hrpf-1(csb26)* double mutant strain compared with *fust-1(csb21)* strain. (B) Confocal images showing the nuclear localization of mRFP-tagged FUST-1 in wildtype N2 strain. DAPI (4',6-diamidino-2-phenylindole) was used to stain nuclei. Note the colocalization of mRFP and DAPI signal. Worm stage: L1. A: Anterior, D: Dorsal. Scale bar: 20 µm. (C) Confocal images showing mRFP-tagged FUST-1 in *fust-1(csb21)* strain. The same worm as in Figure 3.13D with a different focal plane. White arrows indicate the nuclei of intestinal cells. Scale bar: 20 µm. (D) RT-qPCR quantification of *fust-1* mRNA levels at the L1 stage of indicated strains. (A, D) Levels are normalized to the N2 strain using *pmp-3* as the reference gene. Results are shown as mean ± sd of three biological replicates. n.d.: not detected. One-way ANOVA, Tukey's multiple comparisons. *p<0.05, **p < 0.01; ns, not significant.

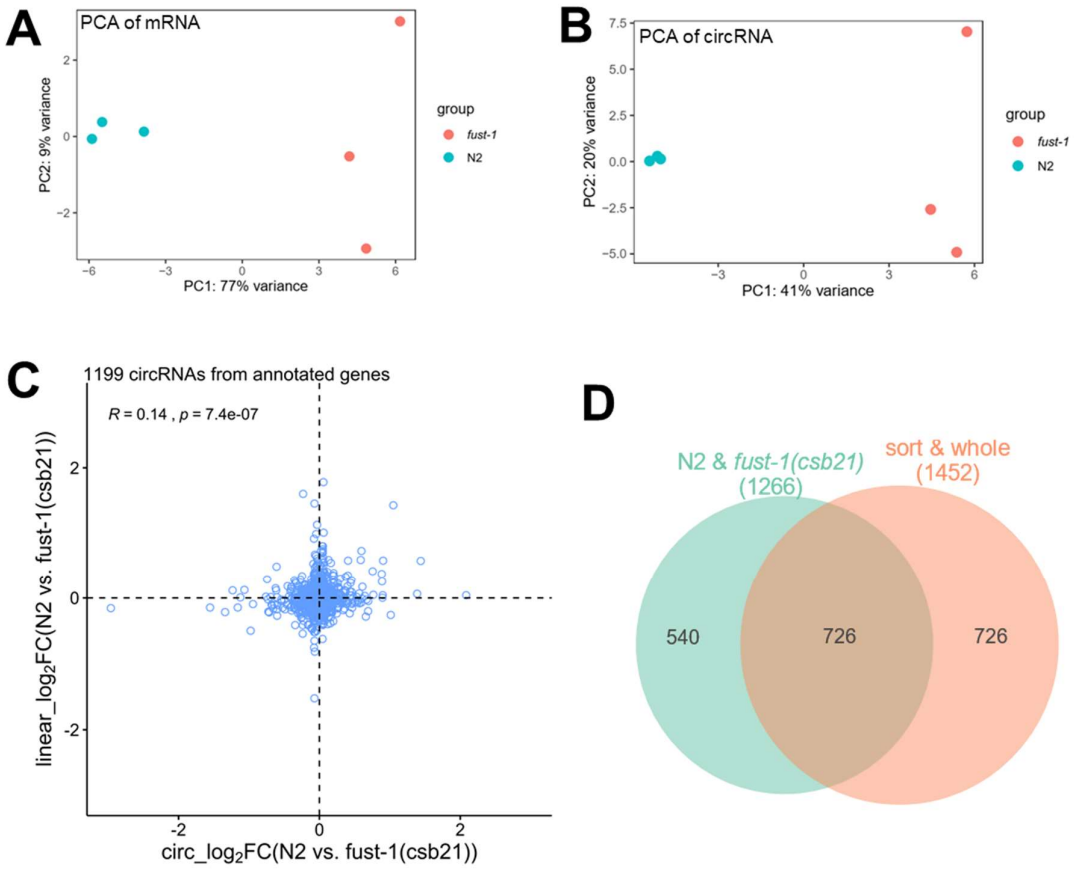


Figure 5.18 Bioinformatic analysis of circRNAs between N2 and *fust-1(csb21)*. Modified from (*161*). (A, B) Principal component analysis (PCA) of linear mRNAs (A) and circRNAs (B) between wildtype N2 strain and *fust-1(csb21)* strain. (C) Scatter plot showing the log2 fold changes of all circRNAs versus log2 fold changes of their cognate linear mRNAs. The Pearson correlation coefficient (*R*) and *p* value (*p*) are shown. (D) Venn diagram showing overlapped circRNAs between the "N2 & *fust-1(csb21)*" dataset and the "sort & whole" dataset.

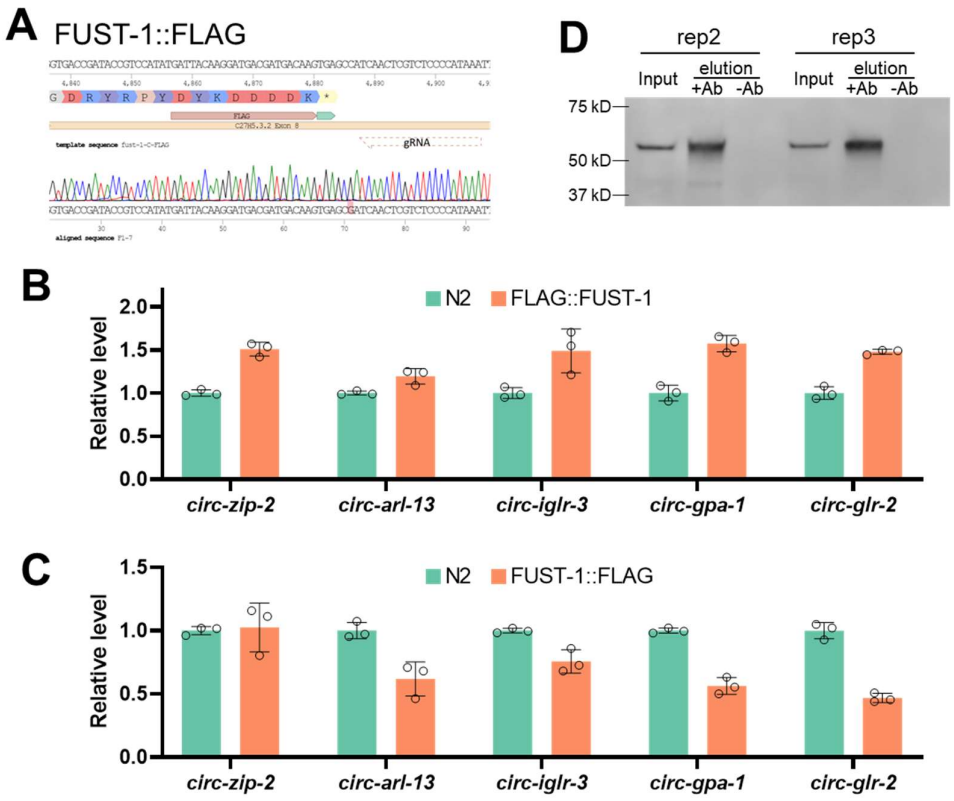


Figure 5.19 FLAG::FUST-1 and FUST-1::FLAG evaluation. Modified from (*161*). (A) Sequence confirmation of C-terminal fusion of FLAG tag just before the stop codon of FUST-1. Note the position of gRNA and the mutated PAM site (TGG>TCG). (B, C) RT-qPCR quantification of circRNA levels between wildtype N2 strain and N-terminal FLAG fusion of FUST-1 strain (B) and between wildtype N2 strain and C-terminal FLAG fusion of FUST-1 strain (C). Levels are normalized to the N2 strain using *pmp-3* as the reference gene. Results are shown as mean ± sd of three biological replicates. (D) Western blot of the other two biological replicates of FLAG::FUST-1 Co-IP with or without FLAG antibody.

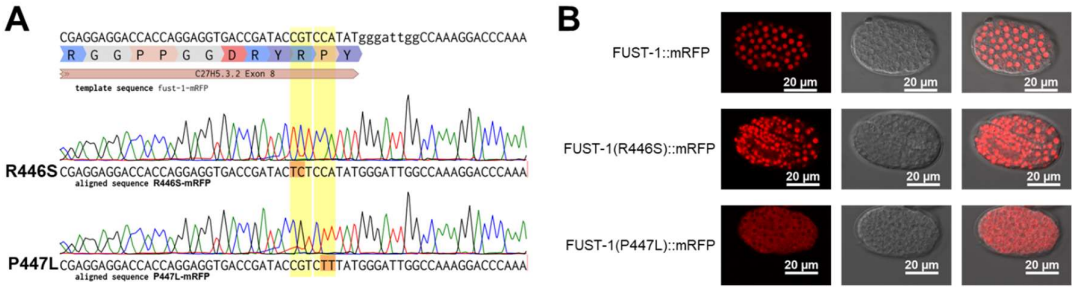


Figure 5.20 Sequence confirmation and expression pattern of FUST-1 knock-in mutant strains. (A) Sequence confirmation of the two knock-in strains. (B) Representative confocal images showing the expression patterns of mRFP-tagged wild-type FUST-1 and mutated FUST-1. Note the absence of mRFP signal in the nuclei in P447L mutation. Scale bars: 20 μm.

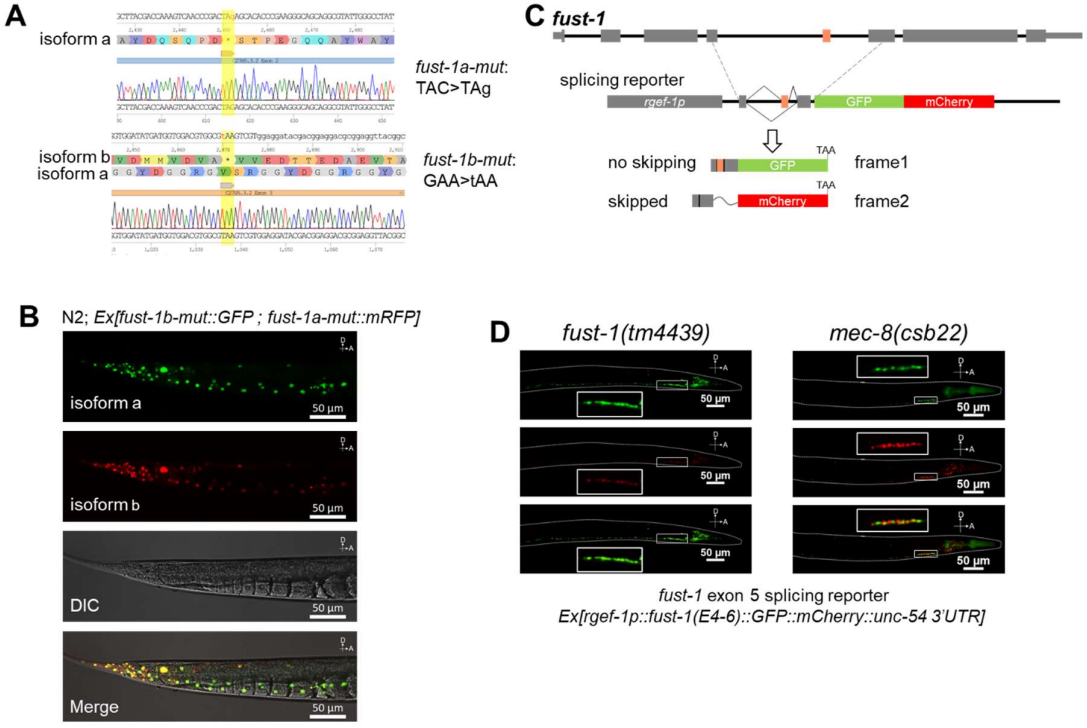


Figure 5.21 Alternative splicing of *fust-1* exon 5. Modified from (*161*). (A) Sequence confirmation of *fust-1a-mut::mRFP* plasmid and *fust-1b-mut::GFP* plasmid. The mutated sites are in yellow shadows. Note the introduction of the stop codons in the read frame of isoform a and isoform b. (B) Confocal images showing expression of FUST-1 isoform a and isoform b in the tail. (C) Illustration of the dual-color splicing reporter of *fust-1* exon 5. The *rgef-1* promoter is used for pan-neuronal expression. (D) Representative images showing the expression patterns of splicing reporter of *fust-1* exon 5 in *fust-1(csb21)* strain and *fust-1(tm4439)* strain. Inset squares show the enlarged neck neurons. (B, D) Worm stage: young adult. A: Anterior, D: Dorsal. Scale bars: 50 μm.

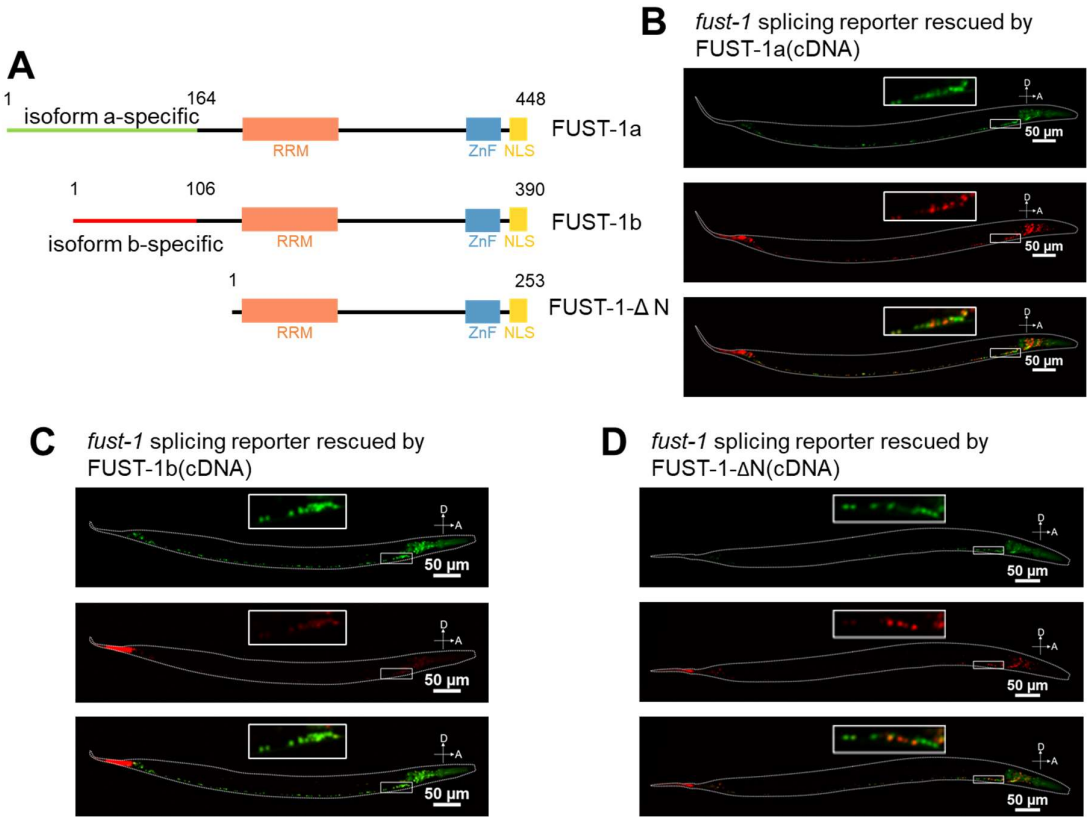


Figure 5.22 Rescue of the alternative splicing reporter of *fust-1* exon 5. Modified from (*161*). (A) Domains in three FUST-1 constructs. The lengths of the constructs are labeled. (B, C, and D) Representative confocal images showing the rescue of *fust-1* exon 5 splicing reporter in the *fust-1(csb21)* strain by FUST-1a (B), FUST-1b (C), and by FUST-1-ΔN (D). Worms were at the young adult stage. Inset squares show the enlarged neck neurons in indicated strains. A: Anterior, D: Dorsal. Scale bars: 50 μm.

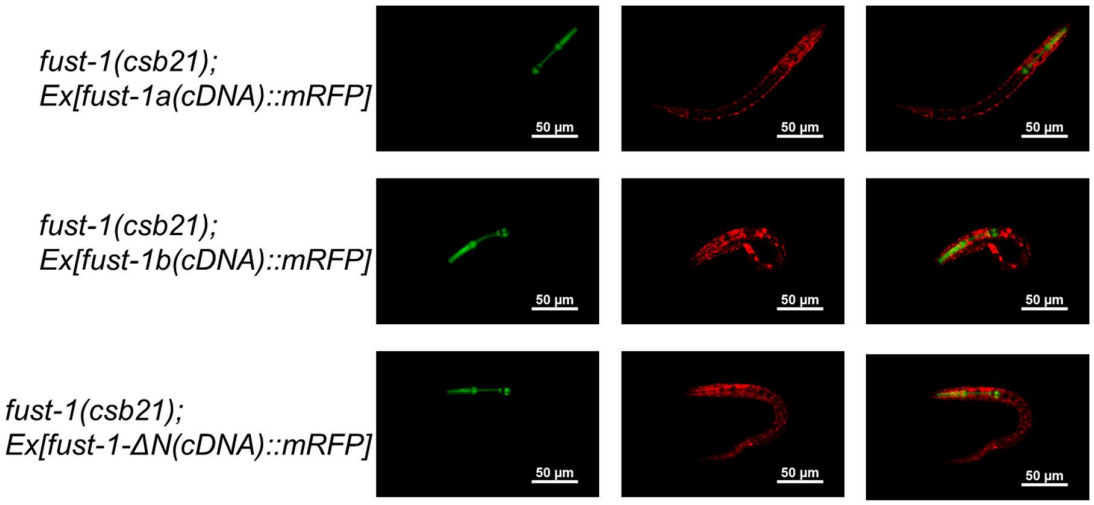


Figure 5.23 Expression patterns of FUST-1 isoforms in the *fust-1(csb21)* strain. Modified from (*161*). Representative confocal images of rescue strains with extrachromosomal expressions of mRFP fused cDNA of FUST-1 isoforms in the *fust-1(csb21)* strain. Note the expression of pharyngeal GFP. Worms were at the L1 stage. Scale bars: 50 μm.

Table 5.9 *C. elegans* strains used in this thesis.

Strains	Genotype	Source	Note
NW1229	*dpy-20(e1362) IV*; *evIs111 [F25B3.3::GFP + dpy-20(+)]*	CGC	Pan-neuronal GFP expression
OF1387	*arl-13(ix262)*	This work	*arl-13* (RCM)
OF1389	*glr-2(ix264)*	This work	*glr-2* (RCM)
OF1390	*gpa-1(ix265)*	This work	*gpa-1* (RCM)
OF1391	*unc-75(ix266)*	This work	*unc-75* RCM deletion #1
OF1392	unc-75(ix267)	This work	*unc-75* RCM deletion #2
OF1393	*iglr-3(ix268)*	This work	*iglr-3* (RCM)
OF1394	*Y20F4.4(ix269)*	This work	*Y20F4.4* (RCM)
OF1395	*zip-2(ix270)*	This work	Deletion of whole upstream intron (intron 1)
OF1471	*gpa-1(ix265); zip-2(ix270)*	This work	Double mutation
OF1472	*Y20F4.4(ix269); zip-2(ix270)*	This work	Double mutation
RM2005	*unc-75(md1344)*	CGC	
CB950	*unc-75(e950)*	CGC	
0F1440	*zip-2(ix310)*	This work	*zip-2*, RCM deletion in intron 1
OF1441	*zip-2(ix311)*	This work	*zip-2*, RCM deletion in intron 4
OF1396	*zip-2(ix271)*	This work	*zip-2* (skip-5'ss)
OF1401	*zip-2(ix275)*	This work	*zip-2* (skip-3'ss)
OF1449	*zip-2(ix313)*	This work	*zip-2* (skip-BP)
OF1399	*zip-2(ix274)*	This work	*zip-2* (circ-5'ss)
OF1398	*zip-2(ix273)*	This work	*zip-2* (circ-3'ss)
OF1451	*zip-2(ix315)*	This work	*zip-2* (circ-BP)
	unc-75(csb20)	Adam Norris	*myo-2::GFP*
	fust-1(csb21)	Adam Norris	*myo-2::GFP*
	mec-8(csb22)	Adam Norris	*myo-2::GFP*
	msi-1(csb24)	Adam Norris	*myo-2::GFP*
	hrpf-1(csb26)	Adam Norris	*myo-2::GFP*
	exc-7(csb29)	Adam Norris	*myo-3::GFP*
	mbl-1(csb30)	Adam Norris	*myo-2::GFP*
	asd-1(csb32)	Adam Norris	*myo-2::GFP*
	tiar-3(csb35)	Adam Norris	*myo-3::GFP*
	Y57G11C.9(csb36)	Adam Norris	*myo-3::GFP*
	tdp-1(csb37)	Adam Norris	*myo-2::GFP*
	fox-1(csb39)	Adam Norris	*myo-2::GFP*

	C25A1.4(csb40)	Adam Norris	*myo-2::GFP*
	fust-1(tm4439)	Alex Parker	
OF1401	*fust-1(csb21); ixEx276[fust-1p::fust-1::mRFP]*	This work	*fust-1* rescue strain (genomic sequences)
OF1402	N2; *ixEx277[fust-1p::fust-1::mRFP]*	This work	*fust-1* overexpression strain (genomic sequences)
OF1434	*fust-1(csb21); hrpf-1(csb26)*	This work	double mutant
OF1405	*fust-1(ixIs280)*	This work	FUST-1::FLAG (C terminal FLAG insertion)
OF1406	*fust-1(ixIs281)*	This work	FLAG::FUST-1 (N terminal FLAG insertion)
OF1435	*zip-2(ix271); fust-1(csb21)*	This work	*zip-2* (skip-5'ss) crossed with *fust-1(csb21)*
OF1436	*zip-2(ix273); fust-1(csb21)*	This work	*zip-2* (circ-3'ss) crossed with *fust-1(csb21)*
OF1437	*zip-2(ix274); fust-1(csb21)*	This work	*zip-2* (circ-5'ss) crossed with *fust-1(csb21)*
OF1438	*zip-2(ix275); fust-1(csb21)*	This work	*zip-2* (skip-3'ss) crossed with *fust-1(csb21)*
OF1453	*zip-2(ix315); fust-1(csb21)*	This work	*zip-2* (circ-BP) crossed with *fust-1(csb21)*
OF1454	*zip-2(ix313); fust-1(csb21)*	This work	*zip-2* (skip-BP) crossed with *fust-1(csb21)*
OF1455	N2; *ixEx[fust-1p::fust-1a-mut::mRFP; fust-1p::fust-1b-mut::GFP]*	This work	expression of FUST-1, 2 isoforms in 2 colors
OF1416	N2; *ixEx291[rgef-1p::fust-1(E4-6)::GFP::mCherry::unc-54 3'UTR]*	This work	*fust-1* exon 5 splicing reporter in N2
OF1417	*fust-1(csb21); ixEx292[rgef-1p::fust-1(E4-6)::GFP::mCherry::unc-54 3'UTR]*	This work	*fust-1* exon 5 splicing reporter in *fust-1(csb21)*
OF1418	*fust-1(tm4439); ixEx293[rgef-1p::fust-1(E4-6)::GFP::mCherry::unc-54 3'UTR]*	This work	*fust-1* exon 5 splicing reporter in *fust-1(tm4439)*
OF1419	*mec-8(csb22); ixEx294[rgef-1p::fust-1(E4-6)::GFP::mCherry::unc-54 3'UTR]*	This work	*fust-1* exon 5 splicing reporter in *mec-8(csb22)*
OF1424	*fust-1(csb21); ixEx299[rgef-1p::fust-1(E4-6)::GFP::mCherry::unc-54 3'UTR; fust-1p::fust-1a(cDNA)::unc-54 3'UTR; lin44p::mRFP]*	This work	Rescue of *fust-1* exon 5 splicing reporter by FUST-1, isoform a cDNA
OF1425	*fust-1(csb21); ixEx300[rgef-1p::fust-1(E4-6)::GFP::mCherry::unc-54 3'UTR; fust-1p::fust-1b(cDNA)::unc-54 3'UTR; lin44p::mRFP]*	This work	Rescue of *fust-1* exon 5 splicing reporter by FUST-1, isoform b cDNA
OF1426	*fust-1(csb21); ixEx301[rgef-1p::fust-1(E4-*	This work	Rescue of *fust-1* exon 5 splicing reporter by FUST-1, ΔN cDNA

	6)::GFP::mCherry::unc-54 3'UTR; fust-1p::fust-1-ΔN(cDNA)::unc-54 3'UTR; lin44p::mRFP]		
OF1408	*fust-1(csb21); ixEx283[fust-1p::fust-1a(cDNA)::mRFP]*	This work	circRNA level rescue by mRFP fused FUST-1, isoform a cDNA
OF1409	*fust-1(csb21); ixEx284[fust-1p::fust-1b(cDNA)::mRFP]*	This work	circRNA level rescue by mRFP fused FUST-1, isoform b cDNA
OF1410	*fust-1(csb21); ixEx285[fust-1p::fust-1a(ΔN)::mRFP]*	This work	circRNA level rescue by mRFP fused FUST-1, ΔN cDNA
OF1407	*ixIs282[fust-1::mRFP]*	This work	FUST-1::mRFP (C-terminal mRFP insertion of FUST-1)
OF1429	*ixIs304[fust-1a::mRFP]*	This work	FUST-1a::mRFP (Only FUST-1, isform a is expressed)

Table 5.10 Primers used in this thesis.

Target	Strand	Sequences (5'-3')
pmp-3	Forward	TGGCCGGATGATGGTGTCGC
	Reverse	ACGAACAATGCCAAAGGCCAGC
circ-nab-1	Forward	GAGGCCTGACGTGGATTTAA
	Reverse	GATGATTGGATCCGATCGAG
circ-unc-75	Forward	CAAAGGCTGCGCATTTCTCA
	Reverse	TCTCTTCCAAGTTTCGGGGG
circ-shw-1	Forward	AGTTTGGGTGGGAAGATGACT
	Reverse	GGATACCGCCTACGTTCAGG
circ-crh-1	Forward	GTCCAATCAGCCACTCGTCT
	Reverse	TGTTGCTGCAGAGGTAGGTG
circ-Y47H9A.1	Forward	CAGTGGAAAAAGTGGTGATCCG
	Reverse	TTTCCTCCTCCGTTGAGTGTC
circ-pig-1	Forward	GCGACACAGGAGACTGTTCA
	Reverse	CCGTTCTGGAACCACTTGGA
circ-gpa-1	Forward	GCGGACAAAGGAGTTCAGTG
	Reverse	CAGCACGGTACTTTTTCCACTC
circ-dmsr-3	Forward	TTGCACAGCTCTCTAGCGTC
	Reverse	AGGAGAAGAAGAACGGCTGAC
circ-mgl-3	Forward	ATTCTGCACAAAGACCGAGG
	Reverse	GGCACGCAAAAATGAAGAGT
circ-Y17G7B.22	Forward	CTGGCGTTGCCCTCATAGAT
	Reverse	TCAGCGTGAAAGGCACTCAT
circ-sma-9	Forward	AACGTGAAAACAGTCCCGAC
	Reverse	GACCACTAGTCAAGCTGCCC
circ-Y20F4.4	Forward	CGACGCCTGGAACAAGAGAA
	Reverse	CGATTAGAAGTGCAGCACCG
circ-glr-2	Forward	TCATGGAACCACTTGGAATGAC
	Reverse	TTGATCTTCATGAGGACGGCA
circ-cam-1	Forward	AGCCTCGGTTAACACGACAA
	Reverse	TTGAGTGGTGGGGTTCCAAG
circ-marc-4	Forward	TCGCGTACACCTGTCTCATC
	Reverse	CGAGTCGACATTTCCAGCCA
circ-iglr-3	Forward	CTGTACGGCCTCGAATGCTT
	Reverse	TGAAGTGTCAACGAGGCACA
circ-arl-13	Forward	ACGGATGAGACATTCACCGAG
	Reverse	GAGAACCCGTTGGTACGGAG
circ-unc-68	Forward	ATGAAAACTGGAAAGGCGGC
	Reverse	GAGCACCGGAACGATCATTTG
circ-ctbp-1	Forward	CGTACTCAAAGAAGAGCCGG
	Reverse	CACGATACCGAGAGCGAAAT

circ-zip-2	Forward	ACATTTCTCCTCCAGCGTCG
	Reverse	AGAAATCTTCGGGAAGGCCG
L-iglr-3	Forward	CTGTACGGCCTCGAATGCTT
	Reverse	TCTCGGCCCAAGTGACAAAA
L-gpa-1	Forward	TTACTGAGTTGTGGGCGGAC
	Reverse	TTGGGAGTGTGAACCCTTGC
L-unc-75	Forward	GCTTTTTGTCGGGCAGATCC
	Reverse	AGAAATGCGCAGCCTTTGTG
L-arl-13	Forward	GGCGGGGATAAGGGAATACG
	Reverse	GAGAGCCTCAATCGACTCGG
L-zip-2#1	Forward	AGGGACCAGTTTCAAGGTCC
	Reverse	AGAAATCTTCGGGAAGGCCG
L-zip-2#2	Forward	ACATTTCTCCTCCAGCGTCG
	Reverse	AGTTTTGCGCCCACGTGTTA
L-glr-2	Forward	CCCAGGAGAGAGAACGAGCA
	Reverse	CCAGGACACGAGGAAAATCGT
L-Y20F4.4	Forward	CGACGCCTGGAACAAGAGAA
	Reverse	GTCTCGCGTAGCTTGTCGAT
zip-2-skip	Forward	TCGCCTTCTCCAGTTCAATCA
	Reverse	TTTTGGCTTGGCGCTTTTGA
arl-13-skip	Forward	CGTAGAAAGTGAGAAGAAATGACCG
	Reverse	GGAGTGGAGAGCCGTCTGAT
Y20F4.4-skip	Forward	TATTCGTCCACAACCCGTCC
	Reverse	GAGGAGAGTTCACGTGCAAT
Y20F4.4-NB-probe	Forward	GCGGTGCTGCACTTCTAATC
	Reverse	ATTTGTGGAGGTCGACGGTT
act-1-NB-probe	Forward	GAGGTTGCCGCTCTTGTTGTA
	Reverse	AAGGTGTGATGCCAGATCTTCTC
zip-2-NB-probe	Forward	GACGGCCTTCCCGAAGATTT
	Reverse	AATAACGGTCCGCTTTCGGT
fust-1	Forward	ATGAAGCCTACATCGCCGAC
	Reverse	TTTTGGTTCTCCGGTGTTGC
zip-2-skip	Forward	TCGCCTTCTCCAGTTCAATCA
	Reverse	TTTTGGCTTGGCGCTTTTGA
arl-13-skip	Forward	CGTAGAAAGTGAGAAGAAATGACCG
	Reverse	GGAGTGGAGAGCCGTCTGAT
zip-2-pre-mRNA	Forward	TCATCGCCTTCTCCAGgtttg
	Reverse	gatcgccacgtcatcgtcata
gpa-1-pre-mRNA	Forward	ACGCGGCGAACATTTTTCTT
	Reverse	GCTACAGCTGTGGTGCTTTC
arl-13-pre-mRNA	Forward	GCAAGTTTCCGCATCTGAGC
	Reverse	CCGATGCCAAAACATCCGAG
18S rRNA	Forward	ACGAGATTGAGCGATAACA

	Reverse	CTGACTCCACCAGTGTAG
26S rRNA	Forward	TGAACTCAGTCGTGATTACC
	Reverse	CACTCGCCGTTACTAAGG
fust-1-pre-mRNA	Forward	ATGGCCCGGTTTCCTTCAAA
	Reverse	ACGCTCGAAGTCGTTTGGTA

Table 5.11 Guide RNA sequences used for CRISPR-Cas9 mutagenesis.

Target gene	crRNA1 (5'-3')	crRNA2 (5'-3') (optional)	Recombinant oligo (5'-3')	validation primers (5'-3') & enzyme
gpa-1 (RCM)	gccgataccagaatgaaatg	ttgatagcattgatgcacca	acatttttcttacaaataggtggtgacaacctcatCCATGGAAATGCGATTACTAATCAAACATCATTCG	AACGCGGCGAACATTTTTCT + TTTCCACTCTCTCCAGCTCCT
iglr-3 (RCM)	acggcaaatcagcaaattgg	AAGGTGGAGCTTTCTCATTG	tgccgaatatccggaaaaacggcaaatcagcaaatTTGAGGCTAAAATTTTTAGATTTTTCAAGATTTTC	TAGGTCCTGGAACGAACACA + AGCCAGCTTCTGATCCAAGT
unc-75 (RCM)	GACGCCGGGAGATATCCGCG	TTAAAACGGTTCGAAGTTGG	aaatttgctctttccctcccctctacaccacgcACTTCGAACCGTTTTAAAAATTACAAATGTAATAG	GCTTTTTGTCGGGCAGGTTG + CCAAGTTTCGGGGGATCTGA
arl-13 (RCM)	ACTAATTTCAGGCTTGTCGT	TGCTCAGATGCGGAAACTTG	aaaggcgtaaaactgtaaaattatgatttccgacgGTTTCCGCATCTGAGCAATATTTTTAGTTTGATGA	GACCGAAAAGTCGTGGTTCG + GTCTCGGCACGATCTATGGG
Y20F4.4 (RCM)	GTAAAGTTACGATTTTGGCA	GAAAATATAATTGTTTCTTG	gagagttcacctggaaaatataattgtttcTTGAGCACTTCAGATAGTTAAAGGTGTAGTAGAACCAAAATCGTAACTTTACAAAACTATAAAAACGACC	CAGAAGGGTACGGGTCGTTT + GTCTCGCGTAGCTTGTCGAT
glr-2 (RCM)	tttttgcttcaacggacacg	ggcattatgcacatagtggg	tttcaactaaatctttttgcttcaacggacGGGCGGAGTTAAAAAATATATGTCTAGAAG	TGATAAGTGAGTAGCACGGAAAC + ACGGCAATATAATCCTTTTCTCCCA
zip-2 (intron 1 whole del)	GTATATGACAAACCTGGAGA	AGCTAGCTTCTAGCTAGGTG	aaacgtccatcagttcgctttcatcgccttctccaAAAAGTTTTCCCCACCTAGCTAGAAGCTAGCTCGTTTGTCACTAGATTTC	AGGGACCAGTTTCAAGGTCC + AGAAATCTTCGGGAAGGCCG
zip-2 (RCM1)	GTATATGACAAACCTGGAGA	AGCTAGCTTCTAGCTAGGTG	TCAAAACGTCCATCAGTTCGCTTTCATCGCgTTCTCCAGgtttgtcatatacaaataataattaccttcgatcaccaatataaaaaatcaatacaattcaaattaaaacgaataaaatcgataattaatatatttcagAAAAGTTTTCgCCACCTAGCTAGAAGCTAGCTCGTTTGTCACTAGATTTCG	AGGGACCAGTTTCAAGGTCC + AGAAATCTTCGGGAAGGCCG
zip-2 (RCM2)	TTATTATTGTGCACAAGCTG	TGTCGTTTCAACTGTTGAGG	CGGACATTGATCCAGTTGACGAGTTCTTTCCACAGCTTGTGCACAATAATAAAAGTAAGCTTAAGACAGTATCACACTAATTAGGAAAATTGAACCCATTTATATATTCTTTAACTCAGCTCTTT + TAGGAAAATTGAACCCATTTATATATTCTTTAACTCAGCTCTTTGTTCAATAAACATTAAAAACCATTTCAGTTCAATCATCTTCCTCCTCAACAGTTGAAACGACAATAACACGTGGGCGCAAA	ACATTTCTCCTCCAGCGTCG + AGTTTTGCGCCCACGTGTTA
zip-2 (skip-5'ss)	GTATATGACAAACCTGGAGA		ATCAAAACGTCCATCAGTTCGCTTTCATCGCgTTCTCCAAttgtcatatacaaataataattacctttttaaatc	AGGGACCAGTTTCAAGGTCC + AGAAATCTTCGGGAAGGCCG (MfeI)

zip-2 (skip-3'ss)	TGTCGTTTCAACTGTTGAGG		taactcaCctctttgttcaatttaaacattaaaaaccatttcTTCAATCATCTTaCTCCTCAata-TTGAAACGACAATAACACGTGGGCGCAAAACTTCA	ACATTTCTCCTCCAGCGTCG + AGTTTTGCGCCCACGTGTTA (SspI)
zip-2 (skip-BP)	TGTCGTTTCAACTGTTGAGG		aagcttaagacagtatcacactaattaggaaaatccacgtcatcgtctgGGcccGtttGtGtGttctttGG-ctcGgctctttgttcGatttaaacGttGGGGGccGtttcagTTCAATCATCTTCaTCCTCAACAGTTGAAACGACAATAACACGTGGGCGCAAAACTTC	ACATTTCTCCTCCAGCGTCG + AGTTTTGCGCCCACGTGTTA (TaqI)
zip-2 (circ-5'ss)	TTATTATTGTGCACAAGCTG		CCAGCGTCGGACATTGATCCAGTTGACGAGTTCTTTaCACAGCTTGTGCACAATAATAAAAtagcttaagacagtatcacactaattaggaaaatcc	AGGGACCAGTTTCAAGGTCC + AGAAATCTTCGGGAAGGCCG (DraI)
zip-2 (circ-3'ss)	AGCTAGCTTCTAGCTAGGTG		AAACGAATAAAATCGATAATTAATATAtttAAAActtttcAccacCTAcCTAcAAGCTAcCTCGTTTGTCACTAcATTTCGACGGCCTTCCCGAAGATTT	AGGGACCAGTTTCAAGGTCC + AGAAATCTTCGGGAAGGCCG (DraI)
zip-2 (circ-BP)	AGCTAGCTTCTAGCTAGGTG		AATAATAATTACCTTTTTAAATCATTATCTTATGACGATGACGTGGCGGTCGCCGGTGTGGGGGGTCGGTGCGGTTCGGGTTGGGGCGGGTGGGGTCGGTGGTTGGTGTGTTTCAGAAAAGTTTTCGCCACCTAGCTAGAAGCTAGCTCGTTTGTCACTAGATTTCGACGG	AGGGACCAGTTTCAAGGTCC + AGAAATCTTCGGGAAGGCCG (HpaII)
flag::fust-1	TTTTGTCGATATTCAAGTCG		tcaacgcatcgacacgtagcgagtgGctcgacttgaatatcgacaaaaatgGATTACAAGGATGACGATGACAAGggttagtttctttttaatagtcgt	GGACAGCGTTCTCCGTCTTC + GTCGTAAGCCGCTGAAATCG
fust-1::flag	TTTATGGGGAGACGAGTTGA		ggaccaccaggaggtgaccgataccgtccatatGATTACAAGGATGACGATGACAAGtgagcGatcaactcgtctccccataaattttaacatctaattt	CCCGAGACCAGATGGAGGAT + GATGGGTAGGGAAATGCGGA
fust-1::mRFP	TTTATGGGGAGACGAGTTGA		dsDNA fragment amplified from plasmid *fust-1p::fust-1::mRFP* (GAGGATCCGGAGGTGGTG + agaatttaaattagatgttaaaatttatgggagacgagttgatcgctcaTTAGGCGCCGGTGGAGTGGC)	
fust-1a::mRFP	TTTTGTCGATATTCAAGTCG	TTTATGGGGAGACGAGTTGA	dsDNA fragment amplified from plasmid containing *fust-1a(cDNA)::mRFP* (taaatttcagtcaacgcatcgacacgtagcgagtgGctcgacttgaatatcgacaaaaatg +agaatttaaattagatgttaaaatttatggggagacgagttgatcgctcaTTAGGCGCCGGTGGAGTGGC)	

6. References

1. E. T. Wang *et al.*, Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470-476 (2008).
2. X. Li, L. Yang, L. L. Chen, The Biogenesis, Functions, and Challenges of Circular RNAs. *Mol Cell* **71**, 428-442 (2018).
3. A. C. Panda, I. Grammatikakis, R. Munk, M. Gorospe, K. Abdelmohsen, Emerging roles and context of circular RNAs. *Wiley Interdiscip Rev RNA* **8**, e1386 (2017).
4. Z. Li *et al.*, Exon-intron circular RNAs regulate transcription in the nucleus. *Nat Struct Mol Biol* **22**, 256-264 (2015).
5. Y. Zhang *et al.*, Circular intronic long noncoding RNAs. *Mol Cell* **51**, 792-806 (2013).
6. H. L. Sanger, G. Klotz, D. Riesner, H. J. Gross, A. K. Kleinschmidt, Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc Natl Acad Sci U S A* **73**, 3852-3856 (1976).
7. H. J. Gross *et al.*, Nucleotide sequence and secondary structure of potato spindle tuber viroid. *Nature* **273**, 203-208 (1978).
8. A. C. Arnberg, G. J. B. Vanommen, L. A. Grivell, E. F. J. Vanbruggen, P. Borst, Some Yeast Mitochondrial Rnas Are Circular. *Cell* **19**, 313-319 (1980).
9. P. J. Grabowski, A. J. Zaug, T. R. Cech, The Intervening Sequence of the Ribosomal-Rna Precursor Is Converted to a Circular Rna in Isolated-Nuclei of Tetrahymena. *Cell* **23**, 467-476 (1981).
10. K. Kruger *et al.*, Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **31**, 147-157 (1982).
11. A. Kos, R. Dijkema, A. C. Arnberg, P. H. van der Meide, H. Schellekens, The hepatitis delta (delta) virus possesses a circular RNA. *Nature* **323**, 558-560 (1986).
12. J. M. Nigro *et al.*, Scrambled exons. *Cell* **64**, 607-613 (1991).
13. C. Cocquerelle, P. Daubersies, M. A. Majerus, J. P. Kerckaert, B. Bailleul, Splicing with inverted order of exons occurs proximal to large introns. *EMBO J* **11**, 1095-1098 (1992).
14. C. Cocquerelle, B. Mascrez, D. Hetuin, B. Bailleul, Mis-splicing yields circular RNA molecules. *Faseb J* **7**, 155-160 (1993).
15. B. Capel *et al.*, Circular transcripts of the testis-determining gene Sry in adult mouse testis. *Cell* **73**, 1019-1030 (1993).
16. X. F. Li, J. Lytton, A circularized sodium-calcium exchanger exon 2 transcript. *J Biol Chem* **274**, 8153-8160 (1999).
17. C. W. Chao, D. C. Chan, A. Kuo, P. Leder, The mouse formin (Fmn) gene: abundant circular RNA transcripts and gene-targeted deletion analysis. *Mol Med* **4**, 614-628 (1998).
18. K. T. Lin, P. S. Lai, M. C. Lin, P. Y. K. Wong, Characterization of circular RNA of cytochrome P450 (CYP)2C23 gene transcripts from rat kidney. *Faseb J* **14**, A1431-A1431 (2000).
19. H. Suzuki *et al.*, Characterization of RNase R-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing. *Nucleic Acids Res* **34**, e63 (2006).
20. A. Soma *et al.*, Permuted tRNA genes expressed via a circular RNA intermediate in Cyanidioschyzon merolae. *Science* **318**, 450-453 (2007).

21. J. Salzman, C. Gawad, P. L. Wang, N. Lacayo, P. O. Brown, Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *Plos One* **7**, e30733 (2012).
22. X. O. Zhang *et al.*, Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res* **26**, 1277-1287 (2016).
23. C. Schindewolf, S. Braun, H. Domdey, In vitro generation of a circular exon from a linear pre-mRNA transcript. *Nucleic Acids Res* **24**, 1260-1266 (1996).
24. Z. Pasman, M. D. Been, M. A. Garcia-Blanco, Exon circularization in mammalian nuclear extracts. *RNA* **2**, 603-610 (1996).
25. S. Braun, H. Domdey, K. Wiebauer, Inverse splicing of a discontinuous pre-mRNA intron generates a circular exon in a HeLa cell nuclear extract. *Nucleic Acids Res* **24**, 4152-4157 (1996).
26. B. Bailleul, During in vivo maturation of eukaryotic nuclear mRNA, splicing yields excised exon circles. *Nucleic Acids Res* **24**, 1015-1019 (1996).
27. C. A. Schindewolf, H. Domdey, Splicing of a circular yeast pre-mRNA in vitro. *Nucleic Acids Res* **23**, 1133-1139 (1995).
28. R. Ashwal-Fluss *et al.*, circRNA biogenesis competes with pre-mRNA splicing. *Mol Cell* **56**, 55-66 (2014).
29. S. Starke *et al.*, Exon circularization requires canonical splice signals. *Cell Rep* **10**, 103-111 (2015).
30. Y. Wang, Z. Wang, Efficient backsplicing produces translatable circular mRNAs. *RNA* **21**, 172-179 (2015).
31. D. Liang, J. E. Wilusz, Short intronic repeat sequences facilitate circular RNA production. *Genes Dev* **28**, 2233-2247 (2014).
32. D. Liang *et al.*, The Output of Protein-Coding Genes Shifts to Circular RNAs When the Pre-mRNA Processing Machinery Is Limiting. *Mol Cell* **68**, 940-954 e943 (2017).
33. M. Wang, J. Hou, M. Muller-McNicoll, W. Chen, E. M. Schuman, Long and Repeat-Rich Intronic Sequences Favor Circular RNA Formation under Conditions of Reduced Spliceosome Activity. *iScience* **20**, 237-247 (2019).
34. Y. Zhang *et al.*, The Biogenesis of Nascent Circular RNAs. *Cell Rep* **15**, 611-624 (2016).
35. S. Kelly, C. Greenman, P. R. Cook, A. Papantonis, Exon Skipping Is Correlated with Exon Circularization. *J Mol Biol* **427**, 2414-2417 (2015).
36. S. P. Barrett, P. L. Wang, J. Salzman, Circular RNA biogenesis can proceed through an exon-containing lariat precursor. *Elife* **4**, e07540 (2015).
37. W. R. Jeck *et al.*, Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19**, 141-157 (2013).
38. A. Surono *et al.*, Circular dystrophin RNAs consisting of exons that were skipped by alternative splicing. *Hum Mol Genet* **8**, 493-500 (1999).
39. P. G. Zaphiropoulos, Exon skipping and circular RNA formation in transcripts of the human cytochrome P-450 2C18 gene in epidermis and of the rat androgen binding protein gene in testis. *Mol Cell Biol* **17**, 2985-2993 (1997).
40. P. G. Zaphiropoulos, Circular RNAs from transcripts of the rat cytochrome P450 2C24 gene: correlation with exon skipping. *Proc Natl Acad Sci U S A* **93**, 6536-6541 (1996).
41. M. A. Khan *et al.*, RBM20 Regulates Circular RNA Production From the Titin Gene. *Circ Res* **119**, 996-1003 (2016).
42. X. O. Zhang *et al.*, Complementary sequence-mediated exon circularization. *Cell* **159**, 134-147 (2014).

43. W. R. Jeck, N. E. Sharpless, Detecting and characterizing circular RNAs. *Nat Biotechnol* **32**, 453-461 (2014).
44. L. L. Chen, The expanding regulatory mechanisms and cellular functions of circular RNAs. *Nat Rev Mol Cell Biol* **21**, 475-490 (2020).
45. V. M. Conn *et al.*, A circRNA from SEPALLATA3 regulates splicing of its cognate mRNA through R-loop formation. *Nat Plants* **3**, 17053 (2017).
46. J. Gubbay *et al.*, Inverted repeat structure of the Sry locus in mice. *Proc Natl Acad Sci U S A* **89**, 7953-7957 (1992).
47. R. A. Dubin, M. A. Kazmi, H. Ostrer, Inverted repeats are necessary for circularization of the mouse testis Sry transcript. *Gene* **167**, 245-248 (1995).
48. M. C. Kramer *et al.*, Combinatorial control of Drosophila circular RNA expression by intronic repeats, hnRNPs, and SR proteins. *Genes Dev* **29**, 2168-2182 (2015).
49. A. Ivanov *et al.*, Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep* **10**, 170-177 (2015).
50. M. Cortes-Lopez *et al.*, Global accumulation of circRNAs during aging in *Caenorhabditis elegans*. *BMC Genomics* **19**, 8 (2018).
51. Y. Gao *et al.*, Comprehensive identification of internal structure and alternative splicing events in circular RNAs. *Nat Commun* **7**, 12060 (2016).
52. A. Rybak-Wolf *et al.*, Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. *Mol Cell* **58**, 870-885 (2015).
53. P. Xia *et al.*, A Circular RNA Protects Dormant Hematopoietic Stem Cells from DNA Sensor cGAS-Mediated Exhaustion. *Immunity* **48**, 688-701 e687 (2018).
54. X. Huang *et al.*, Circular RNA circERBB2 promotes gallbladder cancer progression by regulating PA2G4-dependent rDNA transcription. *Mol Cancer* **18**, 166 (2019).
55. X. Gao *et al.*, Circular RNA-encoded oncogenic E-cadherin variant promotes glioblastoma tumorigenicity through activation of EGFR-STAT3 signalling. *Nat Cell Biol* **23**, 278-291 (2021).
56. K. Nishikura, A-to-I editing of coding and non-coding RNAs by ADARs. *Nat Rev Mol Cell Biol* **17**, 83-96 (2016).
57. T. Aktas *et al.*, DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome. *Nature* **544**, 115-119 (2017).
58. V. Pagliarini *et al.*, Sam68 binds Alu-rich introns in SMN and promotes pre-mRNA circularization. *Nucleic Acids Res*, (2019).
59. X. Li *et al.*, Coordinated circRNA Biogenesis and Function with NF90/NF110 in Viral Infection. *Mol Cell* **67**, 214-227 e217 (2017).
60. S. J. Conn *et al.*, The RNA binding protein quaking regulates formation of circRNAs. *Cell* **160**, 1125-1134 (2015).
61. L. Errichelli *et al.*, FUS affects circular RNA expression in murine embryonic stem cell-derived motor neurons. *Nat Commun* **8**, 14741 (2017).
62. T. Fei *et al.*, Genome-wide CRISPR screen identifies HNRNPL as a prostate cancer dependency regulating RNA splicing. *Proc Natl Acad Sci U S A* **114**, E5207-E5215 (2017).
63. L. V. W. Stagsted, E. T. O'Leary, K. K. Ebbesen, T. B. Hansen, The RNA-binding protein SFPQ preserves long-intron splicing and regulates circRNA biogenesis in mammals. *Elife* **10**, (2021).
64. L. Chen *et al.*, Circ-MALAT1 Functions as Both an mRNA Translation Brake and a microRNA Sponge to Promote Self-Renewal of Hepatocellular Cancer Stem Cells. *Adv Sci (Weinh)* **7**, 1900949 (2020).
65. Y. Enuka *et al.*, Circular RNAs are long-lived and display only minimal early alterations in response to a growth factor. *Nucleic Acids Res* **44**, 1370-1383 (2016).

66. H. Gruner, M. Cortes-Lopez, D. A. Cooper, M. Bauer, P. Miura, CircRNA accumulation in the aging mouse brain. *Sci Rep* **6**, 38907 (2016).
67. H. Hall *et al.*, Transcriptome profiling of aging Drosophila photoreceptors reveals gene expression trends that correlate with visual senescence. *BMC Genomics* **18**, 894 (2017).
68. J. O. Westholm *et al.*, Genome-wide analysis of drosophila circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. *Cell Rep* **9**, 1966-1980 (2014).
69. T. B. Hansen *et al.*, miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *EMBO J* **30**, 4414-4422 (2011).
70. C. X. Liu *et al.*, Structure and Degradation of Circular RNAs Regulate PKR Activation in Innate Immunity. *Cell* **177**, 865-880 e821 (2019).
71. O. H. Park *et al.*, Endoribonucleolytic Cleavage of m(6)A-Containing RNAs by RNase P/MRP Complex. *Mol Cell* **74**, 494-507 e498 (2019).
72. R. Jia, M. S. Xiao, Z. Li, G. Shan, C. Huang, Defining an evolutionarily conserved role of GW182 in circular RNA degradation. *Cell Discov* **5**, 45 (2019).
73. S. Niaz, M. U. Hussain, Role of GW182 protein in the cell. *Int J Biochem Cell Biol* **101**, 29-38 (2018).
74. J. W. Fischer, V. F. Busa, Y. Shao, A. K. L. Leung, Structure-Mediated RNA Decay by UPF1 and G3BP1. *Mol Cell* **78**, 70-84 e76 (2020).
75. R. Dong, X. K. Ma, G. W. Li, L. Yang, CIRCpedia v2: An Updated Database for Comprehensive Circular RNA Annotation and Expression Comparison. *Genomics Proteomics Bioinformatics* **16**, 226-233 (2018).
76. A. de Giorgio, J. Krell, V. Harding, J. Stebbing, L. Castellano, Emerging roles of competing endogenous RNAs in cancer: insights from the regulation of PTEN. *Mol Cell Biol* **33**, 3976-3982 (2013).
77. T. B. Hansen *et al.*, Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384-388 (2013).
78. S. Memczak *et al.*, Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333-338 (2013).
79. M. Piwecka *et al.*, Loss of a mammalian circular RNA locus causes miRNA deregulation and affects brain function. *Science* **357**, eaam8526 (2017).
80. J. E. Wilusz, A 360 degrees view of circular RNAs: From biogenesis to functions. *Wiley Interdiscip Rev RNA* **9**, e1478 (2018).
81. B. Kleaveland, C. Y. Shi, J. Stefano, D. P. Bartel, A Network of Noncoding Regulatory RNAs Acts in the Mammalian Brain. *Cell* **174**, 350-362 e317 (2018).
82. Q. Zheng *et al.*, Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. *Nat Commun* **7**, 11215 (2016).
83. C. Y. Yu *et al.*, The circular RNA circBIRC6 participates in the molecular circuitry controlling human pluripotency. *Nat Commun* **8**, 1149 (2017).
84. J. U. Guo, V. Agarwal, H. Guo, D. P. Bartel, Expanded identification and characterization of mammalian circular RNAs. *Genome Biol* **15**, 409 (2014).
85. X. Liu *et al.*, Synthetic Circular RNA Functions as a miR-21 Sponge to Suppress Gastric Carcinoma Cell Proliferation. *Mol Ther Nucleic Acids* **13**, 312-321 (2018).
86. A. Lavenniah *et al.*, Engineered Circular RNA Sponges Act as miRNA Inhibitors to Attenuate Pressure Overload-Induced Cardiac Hypertrophy. *Mol Ther* **28**, 1506-1517 (2020).
87. Z. Wang *et al.*, Synthetic circular multi-miR sponge simultaneously inhibits miR-21 and miR-93 in esophageal carcinoma. *Lab Invest* **99**, 1442-1453 (2019).

88. W. W. Du *et al.*, Foxo3 circular RNA retards cell cycle progression via forming ternary complexes with p21 and CDK2. *Nucleic Acids Res* **44**, 2846-2858 (2016).
89. L. Gao *et al.*, Circular RNAs from BOULE play conserved roles in protection against stress-induced fertility decline. *Sci Adv* **6**, (2020).
90. W. W. Du *et al.*, Foxo3 circular RNA promotes cardiac senescence by modulating multiple factors associated with stress and senescence responses. *Eur Heart J* **38**, 1402-1412 (2017).
91. Y. Zeng *et al.*, A Circular RNA Binds To and Activates AKT Phosphorylation and Nuclear Localization Reducing Apoptosis and Enhancing Cardiac Repair. *Theranostics* **7**, 3842-3855 (2017).
92. W. Bi *et al.*, CircRNA circRNA_102171 promotes papillary thyroid cancer progression through modulating CTNNBIP1-dependent activation of β-catenin pathway. *J Exp Clin Cancer Res* **37**, 275 (2018).
93. W. W. Du *et al.*, A circular RNA circ-DNMT1 enhances breast cancer progression by activating autophagy. *Oncogene* **37**, 5829-5842 (2018).
94. Y. Chen *et al.*, Circular RNA circAGO2 drives cancer progression through facilitating HuR-repressed functions of AGO2-miRNA complexes. *Cell Death Differ* **26**, 1346-1364 (2019).
95. J. Fang *et al.*, A novel circular RNA, circFAT1(e2), inhibits gastric cancer progression by targeting miR-548g in the cytoplasm and interacting with YBX1 in the nucleus. *Cancer Lett* **442**, 222-232 (2019).
96. V. N. S. Garikipati *et al.*, Circular RNA CircFndc3b modulates cardiac repair after myocardial infarction via FUS/VEGF-A axis. *Nat Commun* **10**, 4317 (2019).
97. J. Guarnerio *et al.*, Intragenic antagonistic roles of protein and circRNA in tumorigenesis. *Cell Res* **29**, 628-640 (2019).
98. Q. Li *et al.*, CircACC1 Regulates Assembly and Activation of AMPK Complex under Metabolic Stress. *Cell Metab* **30**, 157-173 e157 (2019).
99. S. Huang *et al.*, Loss of Super-Enhancer-Regulated circRNA Nfix Induces Cardiac Regeneration After Myocardial Infarction in Adult Mice. *Circulation* **139**, 2857-2876 (2019).
100. W. W. Du *et al.*, The Circular RNA circSKA3 Binds Integrin beta1 to Induce Invadopodium Formation Enhancing Breast Cancer Invasion. *Mol Ther* **28**, 1287-1298 (2020).
101. A. K. Hollensen *et al.*, circZNF827 nucleates a transcription inhibitory complex to balance neuronal differentiation. *Elife* **9**, e58478 (2020).
102. C. H. Wong *et al.*, CircFOXK2 Promotes Growth and Metastasis of Pancreatic Ductal Adenocarcinoma by Complexing with RNA-Binding Proteins and Sponging MiR-942. *Cancer Res* **80**, 2138-2149 (2020).
103. L. Wang *et al.*, Circular RNA circRHOT1 promotes hepatocellular carcinoma progression by initiation of NR2F6 expression. *Mol Cancer* **18**, 119 (2019).
104. P. R. Pandey *et al.*, circSamd4 represses myogenic transcriptional activity of PUR proteins. *Nucleic Acids Res* **48**, 3789-3805 (2020).
105. L. Shi *et al.*, A tumor-suppressive circular RNA mediates uncanonical integrin degradation by the proteasome in liver cancer. *Sci Adv* **7**, (2021).
106. D. C. Tatomer, J. E. Wilusz, An Unchartered Journey for Ribosomes: Circumnavigating Circular RNAs to Produce Proteins. *Mol Cell* **66**, 1-2 (2017).
107. T. Schneider, A. Bindereif, Circular RNAs: Coding or noncoding? *Cell Res* **27**, 724-725 (2017).
108. I. Legnini *et al.*, Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. *Mol Cell* **66**, 22-37 e29 (2017).

109. H. Ho-Xuan *et al.*, Comprehensive analysis of translation from overexpressed circular RNAs reveals pervasive translation from linear transcripts. *Nucleic Acids Res* **48**, 10368-10382 (2020).
110. N. R. Pamudurti *et al.*, Translation of CircRNAs. *Mol Cell* **66**, 9-21 e27 (2017).
111. Y. Yang *et al.*, Extensive translation of circular RNAs driven by N(6)-methyladenosine. *Cell Res* **27**, 626-641 (2017).
112. Y. Yang *et al.*, Novel Role of FBXW7 Circular RNA in Repressing Glioma Tumorigenesis. *J Natl Cancer Inst* **110**, (2018).
113. M. Zhang *et al.*, A novel protein encoded by the circular form of the SHPRH gene suppresses glioma tumorigenesis. *Oncogene* **37**, 1805-1814 (2018).
114. M. Zhang *et al.*, A peptide encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma. *Nat Commun* **9**, 4475 (2018).
115. X. Wu *et al.*, A novel protein encoded by circular SMO RNA is essential for Hedgehog signaling activation and glioblastoma tumorigenicity. *Genome Biol* **22**, 33 (2021).
116. X. Chen *et al.*, circRNADb: A comprehensive database for human circular RNAs with protein-coding annotations. *Sci Rep* **6**, 34985 (2016).
117. X. Meng, Q. Chen, P. Zhang, M. Chen, CircPro: an integrated tool for the identification of circRNAs with protein-coding potential. *Bioinformatics* **33**, 3314-3316 (2017).
118. D. B. Dudekula *et al.*, CircInteractome: A web tool for exploring circular RNAs and their interacting proteins and microRNAs. *RNA Biol* **13**, 34-42 (2016).
119. R. A. Wesselhoeft *et al.*, RNA Circularization Diminishes Immunogenicity and Can Extend Translation Duration In Vivo. *Mol Cell* **74**, 508-520 e504 (2019).
120. R. A. Wesselhoeft, P. S. Kowalski, D. G. Anderson, Engineering circular RNA for potent and stable translation in eukaryotic cells. *Nat Commun* **9**, 2629 (2018).
121. L. Qu *et al.*, Circular RNA Vaccines against SARS-CoV-2 and Emerging Variants. *bioRxiv*, 2021.2003.2016.435594 (2021).
122. S. Memczak, P. Papavasileiou, O. Peters, N. Rajewsky, Identification and Characterization of Circular RNAs As a New Class of Putative Biomarkers in Human Blood. *Plos One* **10**, e0141214 (2015).
123. Y. Li *et al.*, Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis. *Cell Res* **25**, 981-984 (2015).
124. Y. Zhang *et al.*, Circular RNAs: emerging cancer biomarkers and targets. *J Exp Clin Cancer Res* **36**, 152 (2017).
125. J. Guarnerio *et al.*, Oncogenic Role of Fusion-circRNAs Derived from Cancer-Associated Chromosomal Translocations. *Cell* **165**, 289-302 (2016).
126. R. Dong *et al.*, CircRNA-derived pseudogenes. *Cell Res* **26**, 747-750 (2016).
127. X. You *et al.*, Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nat Neurosci* **18**, 603-610 (2015).
128. K. Xu *et al.*, Annotation and functional clustering of circRNA expression in rhesus macaque brain during aging. *Cell Discov* **4**, 48 (2018).
129. S. Brenner, The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71-94 (1974).
130. S. Zhang, D. Banerjee, J. R. Kuhn, Isolation and culture of larval cells from *C. elegans*. *Plos One* **6**, e19505 (2011).
131. J. Cheng, F. Metge, C. Dieterich, Specific identification and quantification of circular RNAs from sequencing data. *Bioinformatics* **32**, 1094-1096 (2016).
132. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

133. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
134. D. Angeles-Albores, R. Lee, J. Chan, P. Sternberg, Two new functions in the WormBase Enrichment Suite. *MicroPubl Biol* **2018**, (2018).
135. D. Angeles-Albores, N. L. RY, J. Chan, P. W. Sternberg, Tissue enrichment analysis for *C. elegans* genomics. *BMC Bioinformatics* **17**, 366 (2016).
136. T. Murayama, I. Maruyama, Plate Assay to Determine *Caenorhabditis elegans* Response to Water Soluble and Volatile Chemicals. *Bio-Protocol* **8**, e2740 (2018).
137. K. Kawamura, I. N. Maruyama, Forward Genetic Screen for *Caenorhabditis elegans* Mutants with a Shortened Locomotor Healthspan. *G3 (Bethesda)* **9**, 2415-2423 (2019).
138. C. I. Nussbaum-Krammer, M. F. Neto, R. M. Brielmann, J. S. Pedersen, R. I. Morimoto, Investigating the spreading and toxicity of prion-like proteins using the metazoan model organism *C. elegans*. *J Vis Exp*, 52321 (2015).
139. G. A. Dokshin, K. S. Ghanta, K. M. Piscopo, C. C. Mello, Robust Genome Editing with Short Single-Stranded and Long, Partially Single-Stranded DNA Donors in *Caenorhabditis elegans*. *Genetics* **210**, 781-787 (2018).
140. R. Kaletsky *et al.*, The *C. elegans* adult neuronal IIS/FOXO transcriptome reveals adult phenotype regulators. *Nature* **529**, 92-96 (2016).
141. W. C. Spencer *et al.*, Isolation of specific neurons from *C. elegans* larvae for gene expression profiling. *Plos One* **9**, e112102 (2014).
142. R. M. Fox *et al.*, A gene expression fingerprint of *C. elegans* embryonic motor neurons. *BMC Genomics* **6**, 42 (2005).
143. Y. Zhang *et al.*, Identification of genes expressed in *C. elegans* touch receptor neurons. *Nature* **418**, 331-335 (2002).
144. M. Christensen *et al.*, A primary culture system for functional analysis of *C. elegans* neurons and muscle cells. *Neuron* **33**, 503-514 (2002).
145. S. N. Deffit *et al.*, The *C. elegans* neural editome reveals an ADAR target mRNA required for proper chemotaxis. *Elife* **6**, (2017).
146. S. E. Von Stetina *et al.*, Cell-specific microarray profiling experiments reveal a comprehensive picture of gene expression in the *C. elegans* nervous system. *Genome Biol* **8**, R135 (2007).
147. J. D. Watson *et al.*, Complementary RNA amplification methods enhance microarray identification of transcripts expressed in the *C. elegans* nervous system. *BMC Genomics* **9**, 84 (2008).
148. D. Cao, Reverse complementary matches simultaneously promote both back-splicing and exon-skipping. *BMC Genomics* **22**, 586 (2021).
149. W. Yang, K. Dierking, H. Schulenburg, WormExp: a web-based application for a *Caenorhabditis elegans*-specific gene expression enrichment analysis. *Bioinformatics* **32**, 943-945 (2016).
150. P. Glazar, P. Papavasileiou, N. Rajewsky, circBase: a database for circular RNAs. *RNA* **20**, 1666-1670 (2014).
151. Y. Yang *et al.*, RNA secondary structure in mutually exclusive splicing. *Nat Struct Mol Biol* **18**, 159-168 (2011).
152. G. E. May, S. Olson, C. J. McManus, B. R. Graveley, Competing RNA secondary structures are required for mutually exclusive splicing of the Dscam exon 6 cluster. *RNA* **17**, 222-229 (2011).
153. B. R. Graveley, Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell* **123**, 65-73 (2005).

154. W. Hong, Y. Shi, B. Xu, Y. Jin, RNA secondary structures in Dscam1 mutually exclusive splicing: unique evolutionary signature from the midge. *RNA* **26**, 1086-1093 (2020).
155. Y. Yue *et al.*, Role and convergent evolution of competing RNA secondary structures in mutually exclusive splicing. *RNA Biol* **14**, 1399-1410 (2017).
156. E. Miriami, H. Margalit, R. Sperling, Conserved sequence elements associated with exon skipping. *Nucleic Acids Res* **31**, 1974-1983 (2003).
157. A. M. Zahler, Pre-mRNA splicing and its regulation in *Caenorhabditis elegans*. *WormBook*, 1-21 (2012).
158. R. V. Aroian *et al.*, Splicing in *Caenorhabditis elegans* does not require an AG at the 3' splice acceptor site. *Mol Cell Biol* **13**, 626-637 (1993).
159. L. Herzel, D. S. M. Ottoz, T. Alpert, K. M. Neugebauer, Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat Rev Mol Cell Biol* **18**, 637-650 (2017).
160. W. Chen, E. Schuman, Circular RNAs in Brain and Other Tissues: A Functional Enigma. *Trends Neurosci* **39**, 597-604 (2016).
161. D. Cao, An autoregulation loop in *fust-1* for circular RNA regulation in *Caenorhabditis elegans*. *Genetics*, iyab145 (2021).
162. A. D. Norris, X. Gracida, J. A. Calarco, CRISPR-mediated genetic interaction profiling identifies RNA binding proteins controlling metazoan fitness. *Elife* **6**, e28129 (2017).
163. J. H. Tan, A. G. Fraser, The combinatorial control of alternative splicing in *C. elegans*. *Plos Genet* **13**, e1007033 (2017).
164. B. Rogelj *et al.*, Widespread binding of FUS along nascent RNA regulates alternative splicing in the brain. *Sci Rep* **2**, 603 (2012).
165. S. Ishigaki *et al.*, Position-dependent FUS-RNA interactions regulate alternative splicing events and transcriptions. *Sci Rep* **2**, 529 (2012).
166. D. S. Dichmann, R. M. Harland, fus/TLS orchestrates splicing of developmental regulators during gastrulation. *Genes Dev* **26**, 1351-1363 (2012).
167. B. Chi *et al.*, Interactome analyses revealed that the U1 snRNP machinery overlaps extensively with the RNAP II machinery and contains multiple ALS/SMA-causative proteins. *Sci Rep* **8**, 8755 (2018).
168. Y. Yu, R. Reed, FUS functions in coupling transcription to splicing by mediating an interaction between RNAP II and U1 snRNP. *Proc Natl Acad Sci U S A* **112**, 8608-8613 (2015).
169. D. Jutzi *et al.*, Aberrant interaction of FUS with the U1 snRNA provides a molecular mechanism of FUS induced amyotrophic lateral sclerosis. *Nat Commun* **11**, 6341 (2020).
170. M. Nolan, K. Talbot, O. Ansorge, Pathogenesis of FUS-associated ALS and FTD: insights from rodent models. *Acta Neuropathol Commun* **4**, 99 (2016).
171. M. Hofweber *et al.*, Phase Separation of FUS Is Suppressed by Its Nuclear Import Receptor and Arginine Methylation. *Cell* **173**, 706-719 e713 (2018).
172. S. N. Baskoylu *et al.*, Disrupted Autophagy and Neuronal Dysfunction in *C. elegans* Knock-in Models of FUS Amyotrophic Lateral Sclerosis. *bioRxiv*, 799932 (2019).
173. A. D. Norris *et al.*, A pair of RNA-binding proteins controls networks of splicing events contributing to specialization of neural cell types. *Mol Cell* **54**, 946-959 (2014).
174. M. Thompson *et al.*, Splicing in a single neuron is coordinately controlled by RNA binding proteins and transcription factors. *Elife* **8**, (2019).

175. C. Merritt, G. Seydoux, Transgenic solutions for the germline. *WormBook*, 1-21 (2010).
176. Z. F. Altun, D. H. Hall, in *Worm Atlas*. (2011), chap. Nervous system, general description.
177. A. C. Billi, S. E. Fischer, J. K. Kim, Endogenous RNAi pathways in *C. elegans*. *WormBook*, 1-49 (2014).
178. S. Li *et al.*, Screening for functional circular RNAs using the CRISPR-Cas13 system. *Nat Methods* **18**, 51-59 (2021).
179. Y. Zhang *et al.*, Optimized RNA-targeting CRISPR/Cas13d technology outperforms shRNA in identifying functional circRNAs. *Genome Biol* **22**, 41 (2021).
180. A. Paix, A. Folkmann, D. Rasoloson, G. Seydoux, High Efficiency, Homology-Directed Genome Editing in *Caenorhabditis elegans* Using CRISPR-Cas9 Ribonucleoprotein Complexes. *Genetics* **201**, 47-54 (2015).
181. A. B. Rose, Introns as Gene Regulators: A Brick on the Accelerator. *Front Genet* **9**, 672 (2018).
182. O. Shaul, How introns enhance gene expression. *Int J Biochem Cell Biol* **91**, 145-155 (2017).
183. H. Kuroyanagi, T. Kobayashi, S. Mitani, M. Hagiwara, Transgenic alternative-splicing reporters reveal tissue-specific expression profiles and regulation mechanisms in vivo. *Nat Methods* **3**, 909-915 (2006).
184. H. Kuroyanagi, G. Ohno, H. Sakane, H. Maruoka, M. Hagiwara, Visualization and genetic analysis of alternative splicing regulation in vivo using fluorescence reporters in transgenic *Caenorhabditis elegans*. *Nat Protoc* **5**, 1495-1517 (2010).
185. J. I. Hoell *et al.*, RNA targets of wild-type and mutant FET family proteins. *Nat Struct Mol Biol* **18**, 1428-1431 (2011).
186. Y. Zhou, S. Liu, G. Liu, A. Ozturk, G. G. Hicks, ALS-associated FUS mutations result in compromised FUS alternative splicing and autoregulation. *Plos Genet* **9**, e1003895 (2013).
187. I. Huppertz *et al.*, iCLIP: protein-RNA interactions at nucleotide resolution. *Methods* **65**, 274-287 (2014).
188. E. L. Van Nostrand *et al.*, Robust, Cost-Effective Profiling of RNA Binding Protein Targets with Single-end Enhanced Crosslinking and Immunoprecipitation (seCLIP). *Methods Mol Biol* **1648**, 177-200 (2017).
189. E. L. Van Nostrand *et al.*, Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* **13**, 508-514 (2016).
190. S. M. Markert *et al.*, An ALS-associated mutation in human FUS reduces neurotransmission from *C. elegans* motor neurons to muscles. *bioRxiv*, 860536 (2019).
191. J. Veriepe, L. Fossouo, J. A. Parker, Neurodegeneration in *C. elegans* models of ALS requires TIR-1/Sarm1 immune pathway activation in neurons. *Nat Commun* **6**, 7319 (2015).
192. T. Murakami *et al.*, ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. *Neuron* **88**, 678-690 (2015).
193. A. Vaccaro *et al.*, Mutant TDP-43 and FUS cause age-dependent paralysis and neurodegeneration in *C. elegans*. *Plos One* **7**, e31321 (2012).
194. A. Vaccaro *et al.*, Methylene blue protects against TDP-43 and FUS neuronal toxicity in *C. elegans* and D. rerio. *Plos One* **7**, e42117 (2012).

195. T. Murakami *et al.*, ALS mutations in FUS cause neuronal dysfunction and death in *Caenorhabditis elegans* by a dominant gain-of-function mechanism. *Hum Mol Genet* **21**, 1-9 (2012).
196. Z. B. Zhou *et al.*, Silencing of circRNA.2837 Plays a Protective Role in Sciatic Nerve Injury by Sponging the miR-34 Family via Regulating Neuronal Autophagy. *Mol Ther Nucleic Acids* **12**, 718-729 (2018).
197. D. Bian, Y. Wu, G. Song, Novel circular RNA, hsa_circ_0025039 promotes cell growth, invasion and glucose metabolism in malignant melanoma via the miR-198/CDK4 axis. *Biomed Pharmacother* **108**, 165-176 (2018).
198. S. Zhong *et al.*, Circular RNA hsa_circ_0000993 inhibits metastasis of gastric cancer cells. *Epigenomics* **10**, 1301-1313 (2018).
199. H. D. Zhang *et al.*, Circular RNA hsa_circ_0052112 promotes cell migration and invasion by acting as sponge for miR-125a-5p in breast cancer. *Biomed Pharmacother* **107**, 1342-1353 (2018).
200. Z. Zeng *et al.*, Circular RNA circ-VANGL1 as a competing endogenous RNA contributes to bladder cancer progression by regulating miR-605-3p/VANGL1 pathway. *J Cell Physiol*, (2018).
201. X. Xu, C. Zhang, S. Zheng, The mechanism of circRNA_0046367/miR-33/ABCA1 regulatory system in lipid homeostasis of liver transplant. *Transplantation* **102**, 163-164 (2018).
202. Y. Wang *et al.*, A Zfp609 circular RNA regulates myoblast differentiation by sponging miR-194-5p. *Int J Biol Macromol* **121**, 1308-1313 (2019).
203. H. Wang, Y. Xiao, L. Wu, D. Ma, Comprehensive circular RNA profiling reveals the regulatory role of the circRNA-000911/miR-449a pathway in breast carcinogenesis. *Int J Oncol* **52**, 743-754 (2018).
204. D. H. Qu, B. D. Yan, R. Xin, T. G. Ma, A novel circular RNA hsa_circ_0020123 exerts oncogenic properties through suppression of miR-144 in non-small cell lung cancer. *American Journal of Cancer Research* **8**, 1387-1402 (2018).
205. W. Luan, Y. Shi, Z. Zhou, Y. Xia, J. Wang, circRNA_0084043 promote malignant melanoma progression via miR-153-3p/Snail axis. *Biochem Biophys Res Commun* **502**, 22-29 (2018).
206. X. Dai *et al.*, Exosomal circRNA_100284 from arsenite-transformed cells, via microRNA-217 regulation of EZH2, is involved in the malignant transformation of human hepatic cells by accelerating the cell cycle and promoting cell proliferation. *Cell Death Dis* **9**, 454 (2018).
207. G. Chen *et al.*, Circular RNAs hsa_circ_0032462, hsa_circ_0028173, hsa_circ_0005909 are predicted to promote CADM1 expression by functioning as miRNAs sponge in human osteosarcoma. *Plos One* **13**, e0202896 (2018).
208. D. Chen, W. Ma, Z. Ke, F. Xie, CircRNA hsa_circ_100395 regulates miR-1228/TCF21 pathway to inhibit lung cancer progression. *Cell Cycle* **17**, 2080-2090 (2018).
209. Y. Bai *et al.*, Circular RNA DLGAP4 Ameliorates Ischemic Stroke Outcomes by Targeting miR-143 to Regulate Endothelial-Mesenchymal Transition Associated with Blood-Brain Barrier Integrity. *J Neurosci* **38**, 32-50 (2018).
210. Y. Zhang *et al.*, CircRNA_100269 is downregulated in gastric cancer and suppresses tumor cell growth by targeting miR-630. *Aging (Albany NY)* **9**, 1585-1594 (2017).
211. Y. Yuan, W. Liu, Y. Zhang, Y. Zhang, S. Sun, CircRNA circ_0026344 as a prognostic biomarker suppresses colorectal cancer progression via microRNA-21 and microRNA-31. *Biochem Biophys Res Commun* **503**, 870-875 (2018).

212. H. Xu *et al.*, NFIX Circular RNA Promotes Glioma Progression by Regulating miR-34a-5p via Notch Signaling Pathway. *Front Mol Neurosci* **11**, 225 (2018).
213. Z. Wu *et al.*, Circular RNA CEP128 acts as a sponge of miR-145-5p in promoting the bladder cancer progression via regulating SOX11. *Mol Med* **24**, 40 (2018).
214. L. Wang *et al.*, CircDOCK1 suppresses cell apoptosis via inhibition of miR196a5p by targeting BIRC3 in OSCC. *Oncol Rep* **39**, 951-966 (2018).
215. J. Song *et al.*, CircularRNA_104670 plays a critical role in intervertebral disc degeneration by functioning as a ceRNA. *Exp Mol Med* **50**, 1-12 (2018).
216. X. Cheng *et al.*, Circular RNA VMA21 protects against intervertebral disc degeneration through targeting miR-200c and X linked inhibitor-of-apoptosis protein. *Ann Rheum Dis* **77**, 770-779 (2018).
217. H. Ouyang *et al.*, Circular RNA circSVIL Promotes Myoblast Proliferation and Differentiation by Sponging miR-203 in Chicken. *Front Genet* **9**, 172 (2018).
218. A. Bian *et al.*, Circular RNA Complement Factor H (CFH) Promotes Glioma Progression by Sponging miR-149 and Regulating AKT1. *Med Sci Monit* **24**, 5704-5712 (2018).
219. N. Bai *et al.*, circFBLIM1 act as a ceRNA to promote hepatocellular cancer progression by sponging miR-346. *J Exp Clin Cancer Res* **37**, 172 (2018).
220. X. Zhang *et al.*, circSMAD2 inhibits the epithelial-mesenchymal transition by targeting miR-629 in hepatocellular carcinoma. *Onco Targets Ther* **11**, 2853-2863 (2018).
221. C. Yang *et al.*, Circular RNA circ-ITCH inhibits bladder cancer progression by sponging miR-17/miR-224 and regulating p21, PTEN expression. *Mol Cancer* **17**, 19 (2018).
222. J. Hu *et al.*, The circular RNA circ-ITCH suppresses ovarian carcinoma progression through targeting miR-145/RASA1 signaling. *Biochem Biophys Res Commun* **505**, 222-228 (2018).
223. J. Wu *et al.*, CircIRAK3 sponges miR-3607 to facilitate breast cancer metastasis. *Cancer Lett* **430**, 179-192 (2018).
224. Y. Z. Song, J. F. Li, Circular RNA hsa_circ_0001564 regulates osteosarcoma proliferation and apoptosis by acting miRNA sponge. *Biochem Biophys Res Commun* **495**, 2369-2375 (2018).
225. H. B. Ma, Y. N. Yao, J. J. Yu, X. X. Chen, H. F. Li, Extensive profiling of circular RNAs and the potential regulatory role of circRNA-000284 in cell proliferation and invasion of cervical cancer via sponging miR-506. *Am J Transl Res* **10**, 592-604 (2018).
226. J. Liu, D. Wang, Z. Long, J. Liu, W. Li, CircRNA8924 Promotes Cervical Cancer Cell Proliferation, Migration and Invasion by Competitively Binding to MiR-518d-5p /519-5p Family and Modulating the Expression of CBX8. *Cell Physiol Biochem* **48**, 173-184 (2018).
227. W. J. Huang *et al.*, Silencing circular RNA hsa_circ_0000977 suppresses pancreatic ductal adenocarcinoma progression by stimulating miR-874-3p and inhibiting PLK1 expression. *Cancer Lett* **422**, 70-80 (2018).
228. X. Y. Guo *et al.*, circRNA_0046366 inhibits hepatocellular steatosis by normalization of PPAR signaling. *World J Gastroenterol* **24**, 323-337 (2018).
229. B. Chen *et al.*, circEPSTI1 as a Prognostic Marker and Mediator of Triple-Negative Breast Cancer Progression. *Theranostics* **8**, 4003-4015 (2018).
230. R. He *et al.*, circGFRA1 and GFRA1 act as ceRNAs in triple negative breast cancer by regulating miR-34a. *J Exp Clin Cancer Res* **36**, 145 (2017).

231. B. Zhou, J. W. Yu, A novel identified circular RNA, circRNA_010567, promotes myocardial fibrosis via suppressing miR-141 by targeting TGF-beta1. *Biochem Biophys Res Commun* **487**, 769-775 (2017).
232. C. M. Tang *et al.*, CircRNA_000203 enhances the expression of fibrosis-associated genes by derepressing targets of miR-26b-5p, Col1a2 and CTGF, in cardiac fibroblasts. *Sci Rep* **7**, 40342 (2017).
233. L. Peng *et al.*, Circular RNA ZNF609 functions as a competitive endogenous RNA to regulate AKT3 expression by sponging miR-150-5p in Hirschsprung's disease. *Oncotarget* **8**, 808-818 (2017).
234. X. Y. Huang *et al.*, Comprehensive circular RNA profiling reveals the regulatory role of the circRNA-100338/miR-141-3p pathway in hepatitis B-related hepatocellular carcinoma. *Sci Rep* **7**, 5428 (2017).
235. K. Wang *et al.*, Circular RNA mediates cardiomyocyte death via miRNA-dependent upregulation of MTP18 expression. *Cell Death Differ* **24**, 1111-1120 (2017).
236. D. Han *et al.*, Circular RNA circMTO1 acts as the sponge of microRNA-9 to suppress hepatocellular carcinoma progression. *Hepatology* **66**, 1151-1164 (2017).
237. J. Rao, X. Cheng, H. Zhu, L. Liu, L. Wang, Circular RNA-0007874 (circMTO1) reverses chemoresistance to temozolomide by acting as a sponge of microRNA-630 in glioblastoma. *Cell Biol Int*, (2018).
238. Y. Li, B. Wan, L. Liu, L. Zhou, Q. Zeng, Circular RNA circMTO1 suppresses bladder cancer metastasis by sponging miR-221 and inhibiting epithelial-to-mesenchymal transition. *Biochem Biophys Res Commun* **508**, 991-996 (2019).
239. L. Chen *et al.*, circRNA_100290 plays a role in oral cancer by functioning as a sponge of the miR-29 family. *Oncogene* **36**, 4551-4561 (2017).
240. Z. Zhong, M. Lv, J. Chen, Screening differential circular RNA expression profiles reveals the regulatory role of circTCF25-miR-103a-3p/miR-107-CDK6 pathway in bladder carcinoma. *Sci Rep* **6**, 30919 (2016).
241. H. Xie *et al.*, Emerging roles of circRNA_001569 targeting miR-145 in the proliferation and invasion of colorectal cancer. *Oncotarget* **7**, 26680-26691 (2016).
242. S. Cao, G. Wang, J. Wang, C. Li, L. Zhang, Hsa_circ_101280 promotes hepatocellular carcinoma by regulating miR-375/JAK2. *Immunol Cell Biol* **97**, 218-228 (2019).
243. F. Xie *et al.*, Circular RNA BCRC-3 suppresses bladder cancer proliferation through miR-182-5p/p27 axis. *Molecular Cancer* **17**, 144 (2018).
244. C. Song *et al.*, The competing endogenous circular RNA ADAMTS14 suppressed hepatocellular carcinoma progression through regulating microRNA-572/regulator of calcineurin 1. *J Cell Physiol*, (2018).
245. H. Liu *et al.*, Circular RNA YAP1 inhibits the proliferation and invasion of gastric cancer cells by regulating the miR-367-5p/p27 (Kip1) axis. *Mol Cancer* **17**, 151 (2018).
246. H. Liu *et al.*, Circular RNA circUBXN7 represses cell growth and invasion by sponging miR-1247-3p to enhance B4GALT3 expression in bladder cancer. *Aging (Albany NY)* **10**, 2606-2623 (2018).
247. X. Zhang *et al.*, circRNA_104075 stimulates YAP-dependent tumorigenesis through the regulation of HNF4a and may serve as a diagnostic marker in hepatocellular carcinoma. *Cell Death Dis* **9**, 1091 (2018).
248. C. Yang *et al.*, Silencing Circular RNA UVRAG Inhibits Bladder Cancer Growth and Metastasis via Targeting the miR-223/FGFR2 Axis. *Cancer Sci*, (2018).
249. C. Hu *et al.*, Overexpressed circ_0067934 acts as an oncogene to facilitate cervical cancer progression via the miR-545/EIF3C axis. *J Cell Physiol*, (2018).

250. J. Shang *et al.*, CircPAN3 mediates drug resistance in acute myeloid leukemia through the miR-153-5p/miR-183-5p-XIAP axis. *Exp Hematol*, (2018).
251. X. Chen, H. Ouyang, Z. Wang, B. Chen, Q. Nie, A Novel Circular RNA Generated by FGFR2 Gene Promotes Myoblast Proliferation and Differentiation by Sponging miR-133a-5p and miR-29b-1-5p. *Cells* **7**, (2018).
252. J. Zhang *et al.*, Circular RNA_LARP4 inhibits cell proliferation and invasion of gastric cancer by sponging miR-424-5p and regulating LATS1 expression. *Mol Cancer* **16**, 151 (2017).
253. X. Tian *et al.*, CircABCB10 promotes nonsmall cell lung cancer cell proliferation and migration by regulating the miR-1252/FOXR2 axis. *J Cell Biochem* **120**, 3765-3772 (2019).
254. X. Duan, D. Liu, Y. Wang, Z. Chen, Circular RNA hsa_circ_0074362 Promotes Glioma Cell Proliferation, Migration, and Invasion by Attenuating the Inhibition of miR-1236-3p on HOXB7 Expression. *DNA Cell Biol* **37**, 917-924 (2018).
255. Y. Zhang *et al.*, CircDYM ameliorates depressive-like behavior by targeting miR-9 to regulate microglial activation via HSP90 ubiquitination. *Mol Psychiatry* **25**, 1175-1190 (2020).
256. H. Liu *et al.*, Invasion-related circular RNA circFNDC3B inhibits bladder cancer progression through the miR-1178-3p/G3BP2/SRC/FAK axis. *Mol Cancer* **17**, 161 (2018).
257. K. Zeng *et al.*, The pro-metastasis effect of circANKS1B in breast cancer. *Mol Cancer* **17**, 160 (2018).
258. B. Xie *et al.*, CircRNA has_circ_0078710 acts as the sponge of microRNA-31 involved in hepatocellular carcinoma progression. *Gene* **683**, 253-261 (2019).
259. Y. Xiong, J. Zhang, C. Song, CircRNA ZNF609 functions as a competitive endogenous RNA to regulate FOXP4 expression by sponging miR-138-5p in renal carcinoma. *J. Cell. Physiol.*, Ahead of Print (2018).
260. J.-J. Wang *et al.*, Circular RNA-ZNF609 regulates retinal neurodegeneration by acting as miR-615 sponge. *Theranostics* **8**, 3408-3415 (2018).
261. D.-M. Wu *et al.*, Role of circular RNA DLEU2 in human acute myeloid leukemia. *Mol. Cell. Biol.* **38**, e00259-00218/00251-e00259-00218/00219 (2018).
262. R. Wang *et al.*, CircNT5E acts as a sponge of microRNA-422a to promote glioblastoma tumorigenesis. *Cancer Res.* **78**, 4812-4825 (2018).
263. R. Wang *et al.*, EIF4A3-induced circular RNA MMP9 (circMMP9) acts as a sponge of miR-124 and promotes glioblastoma multiforme cell tumorigenesis. *Mol Cancer* **17**, 166 (2018).
264. Z. Li *et al.*, CircPCNXL2 sponges miR-153 to promote the proliferation and invasion of renal cancer cells through upregulating ZEB2. *Cell Cycle*, (2018).
265. C. Jin, A. Wang, L. Liu, G. Wang, G. Li, Hsa_circ_0136666 promotes the proliferation and invasion of colorectal cancer through miR-136/SH2B1 axis. *J. Cell. Physiol.*, Ahead of Print (2018).
266. B. Lei *et al.*, Circular RNA hsa_circ_0076248 promotes oncogenesis of glioma by sponging miR-181a to modulate SIRT1 expression. *J Cell Biochem* **120**, 6698-6708 (2019).
267. G. Liu *et al.*, CircFAT1 sponges miR-375 to promote the expression of Yes-associated protein 1 in osteosarcoma cells. *Mol Cancer* **17**, 170 (2018).
268. C. Yang *et al.*, Silencing circular RNA UVRAG inhibits bladder cancer growth and metastasis by targeting the microRNA-223/fibroblast growth factor receptor 2 axis. *Cancer Sci* **110**, 99-106 (2019).

269. Y. Zhou *et al.*, Circular RNA hsa_circ_0004015 regulates the proliferation, invasion, and TKI drug resistance of non-small cell lung cancer by miR-1183/PDPK1 signaling pathway. *Biochem Biophys Res Commun* **508**, 527-535 (2019).
270. M. Yang *et al.*, Circular RNA circ_0034642 elevates BATF3 expression and promotes cell proliferation and invasion through miR-1205 in glioma. *Biochem Biophys Res Commun* **508**, 980-985 (2019).
271. F. Liu *et al.*, Circular RNA EIF6 (Hsa_circ_0060060) sponges miR-144-3p to promote the cisplatin-resistance of human thyroid carcinoma cells by autophagy regulation. *Aging (Albany NY)* **10**, 3806-3820 (2018).
272. X. Y. Li *et al.*, Circular RNA circPVT1 Promotes Proliferation and Invasion Through Sponging miR-125b and Activating E2F2 Signaling in Non-Small Cell Lung Cancer. *Cell Physiol Biochem* **51**, 2324-2340 (2018).
273. S. Qin *et al.*, Circular RNA PVT1 acts as a competing endogenous RNA for miR-497 in promoting non-small cell lung cancer progression. *Biomed Pharmacother* **111**, 244-250 (2019).
274. Z. Wang *et al.*, Novel circular RNA NF1 acts as a molecular sponge, promoting gastric cancer by absorbing miR-16. *Endocr Relat Cancer*, (2018).
275. I. F. Hall *et al.*, Circ_Lrp6, a Circular RNA Enriched in Vascular Smooth Muscle Cells, Acts as a Sponge Regulating miRNA-145 Function. *Circ Res*, (2018).
276. J. M. Bi *et al.*, Circular RNA circ-ZKSCAN1 inhibits bladder cancer progression through miR-1178-3p/p21 axis and acts as a prognostic factor of recurrence. *Molecular Cancer* **18**, 133 (2019).
277. B. Chen *et al.*, Circular RNA circHIPK3 Promotes the Proliferation and Differentiation of Chicken Myoblast Cells by Sponging miR-30a-3p. *Cells* **8**, (2019).
278. Z. Cheng *et al.*, circTP63 functions as a ceRNA to promote lung squamous cell carcinoma progression by upregulating FOXM1. *Nat Commun* **10**, 3200 (2019).
279. A. Cherubini *et al.*, FOXP1 circular RNA sustains mesenchymal stem cell identity via microRNA inhibition. *Nucleic Acids Res* **47**, 5325-5340 (2019).
280. W. Dong *et al.*, Circular RNA ACVR2A suppresses bladder cancer cells proliferation and metastasis through miR-626/EYA4 axis. *Mol Cancer* **18**, 95 (2019).
281. X. Huang *et al.*, Circular RNA AKT3 upregulates PIK3R1 to enhance cisplatin resistance in gastric cancer via miR-198 suppression. *Mol Cancer* **18**, 71 (2019).
282. H. Li *et al.*, CircRNA CBL.11 suppresses cell proliferation by sponging miR-6778-5p in colorectal cancer. *BMC Cancer* **19**, 826 (2019).
283. C. Liu *et al.*, Targeting pericyte-endothelial cell crosstalk by circular RNA-cPWWP2A inhibition aggravates diabetes-induced microvascular dysfunction. *Proc Natl Acad Sci U S A* **116**, 7455-7464 (2019).
284. W. Liu, J. Zhao, M. Jin, M. Zhou, circRAPGEF5 Contributes to Papillary Thyroid Proliferation and Metastatis by Regulation miR-198/FGFR1. *Mol Ther Nucleic Acids* **14**, 609-616 (2019).
285. Q. Lu *et al.*, Circular RNA circSLC8A1 acts as a sponge of miR-130b/miR-494 in suppressing bladder cancer progression via regulating PTEN. *Molecular Cancer* **18**, 111 (2019).
286. H. W. Su *et al.*, Circular RNA cTFRC acts as the sponge of MicroRNA-107 to promote bladder carcinoma progression. *Molecular Cancer* **18**, 27 (2019).
287. Y. Wu *et al.*, Circular RNA circTADA2A promotes osteosarcoma progression and metastasis by sponging miR-203a-3p and regulating CREB3 expression. *Mol Cancer* **18**, 73 (2019).
288. J. Z. Xu *et al.*, circTADA2As suppress breast cancer progression and metastasis via targeting miR-203a-3p/SOCS3 axis. *Cell Death Dis* **10**, 175 (2019).

289. S. B. Zhang *et al.*, CircAnks1a in the spinal cord regulates hypersensitivity in a rodent model of neuropathic pain. *Nat Commun* **10**, 4119 (2019).
290. X. Zhang *et al.*, Circular RNA circNRIP1 acts as a microRNA-149-5p sponge to promote gastric cancer progression via the AKT1/mTOR pathway. *Mol Cancer* **18**, 20 (2019).
291. Z. L. Zhu *et al.*, Circular RNA circNHSL1 promotes gastric cancer progression through the miR-1306-3p/SIX1/vimentin axis. *Molecular Cancer* **18**, 126 (2019).
292. L. Y. Chen *et al.*, The circular RNA circ-ERBIN promotes growth and metastasis of colorectal cancer by miR-125a-5p and miR-138-5p/4EBP-1 mediated cap-independent HIF-1alpha translation. *Mol Cancer* **19**, 164 (2020).
293. Q. Chen *et al.*, CircRNA cRAPGEF5 inhibits the growth and metastasis of renal cell carcinoma via the miR-27a-3p/TXNIP pathway. *Cancer Lett* **469**, 68-77 (2020).
294. Q. Chen *et al.*, Circular RNA circSnx5 Controls Immunogenicity of Dendritic Cells through the miR-544/SOCS1 Axis and PU.1 Activity Regulation. *Mol Ther* **28**, 2503-2518 (2020).
295. D. Tian *et al.*, Circ-ADAM9 targeting PTEN and ATG7 promotes autophagy and apoptosis of diabetic endothelial progenitor cells by sponging mir-20a-5p. *Cell Death Dis* **11**, 526 (2020).
296. L. Wang *et al.*, Circular RNA circSEMA5A promotes bladder cancer progression by upregulating ENO1 and SEMA5A expression. *Aging (Albany NY)* **12**, 21674-21686 (2020).
297. J. Xu *et al.*, circEYA1 Functions as a Sponge of miR-582-3p to Suppress Cervical Adenocarcinoma Tumorigenesis via Upregulating CXCL14. *Mol Ther Nucleic Acids* **22**, 1176-1190 (2020).
298. Y. Fan *et al.*, CircNR3C2 promotes HRD1-mediated tumor-suppressive effect via sponging miR-513a-3p in triple-negative breast cancer. *Mol Cancer* **20**, 25 (2021).
299. P. Shen *et al.*, CircNEIL3 regulatory loop promotes pancreatic ductal adenocarcinoma progression via miRNA sponging and A-to-I RNA-editing. *Mol Cancer* **20**, 51 (2021).
300. S. Tan *et al.*, circST6GALNAC6 suppresses bladder cancer metastasis by sponging miR-200a-3p to modulate the STMN1/EMT axis. *Cell Death Dis* **12**, 168 (2021).
301. L. Wang *et al.*, Estrogen-induced circRNA, circPGR, functions as a ceRNA to promote estrogen receptor-positive breast cancer cell growth by regulating cell cycle-related genes. *Theranostics* **11**, 1732-1752 (2021).
302. J. X. Zhou *et al.*, CircularRNA circPARP4 promotes glioblastoma progression through sponging miR-125a-5p and regulating FUT4. *American Journal of Cancer Research* **11**, 138-+ (2021).

7. Appendices

7.1 Sequence alignments

arl-13(ix262)

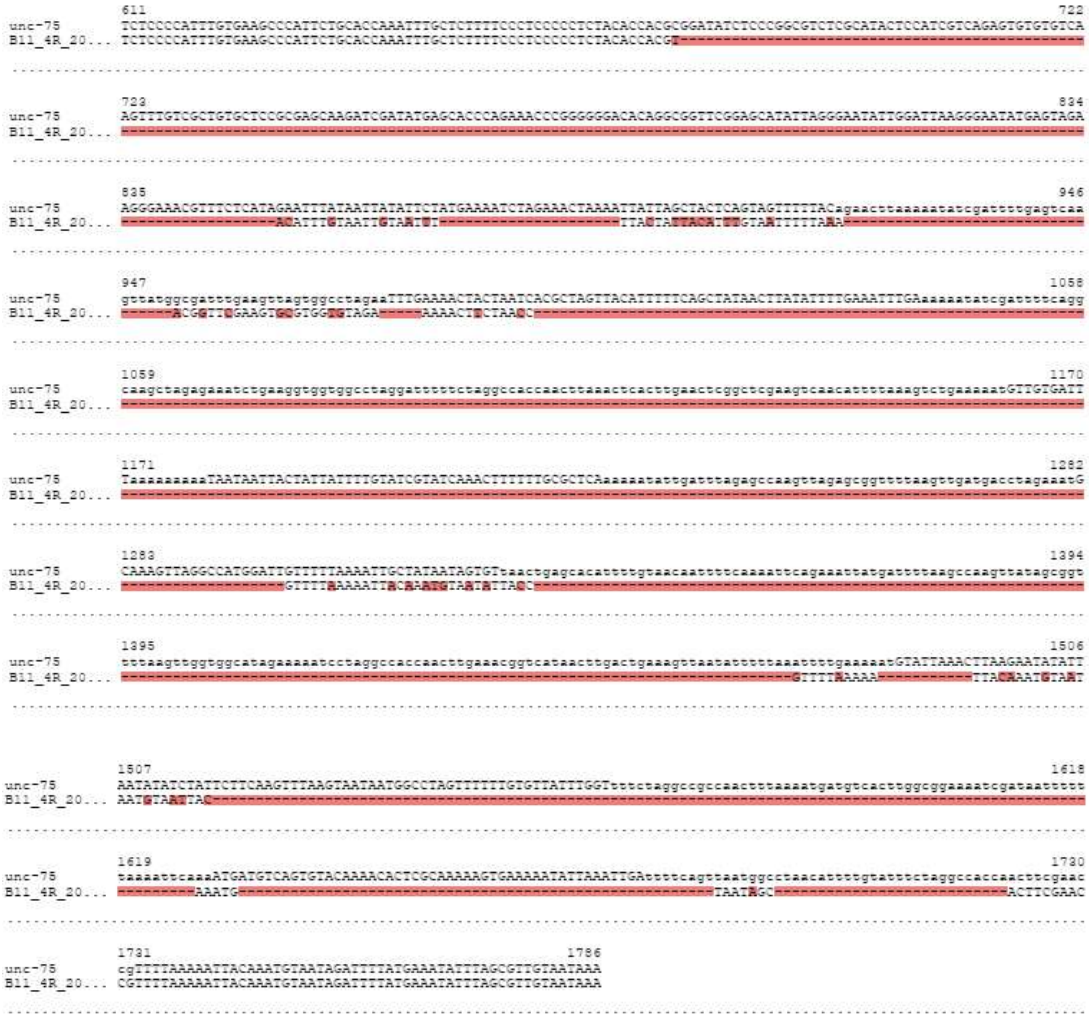
glr-2(ix264)


2237 2318
glr-2b un... agcacggaaacattttataaaattaatgtcaacacagttgaactttcaactaaatctttttgcttcaacggacacgaggaaa
F7-R AGCACGGAAACATTTTATAAAATTAATGTCAACACAGTTGAACTTTCAACTAAATCTTTTTGCTTCAACGGAC---------
2319 2400
glr-2b un... atcgttagagagacctgaaatggtctaataatttttaagattacagtaatacagtgcatttttttccacttttcatgataaaaa
F7-R ----------
2401 2482
glr-2b un... accttcataaagggtttttatgcttctctcttatttttgctcccacggataccccgaagaagcttgtcaaaatttggtgttt
F7-R ----------
2483 2564
glr-2b un... tcgacacactcttgcgaatgatcgaattccatactaaccttggttatatgggttaaaaccaaagctgtaaacgaataaggcg
F7-R ----------
2565 2646
glr-2b un... cacatgactccttgcaccagcggacaactcgcgggtttcgcacaccactctgcattgcatagagactgggcggggctaggag
F7-R ----------
2647 2728
glr-2b un... aacaactactgtagaggctgaaaaaatttaaggcattatgcacatagtgggcggagttaaaaaatatatgtctagaagctta
F7-R ----------GGGCGGAGTTAAAAAATATATGTCTAGAAGCTTA
2729 2810
glr-2b un... tttatatttatatttctgcgaaaaattaaagattttaaaaactgcataaaaaatcatccatatatttagATCATTTGTACGT
F7-R TTTATATTTATATTTCTGCGAAAAATTAAAGATTTTAAAAACTGCATAAAAAATCATCCATATATTTAGATCATTTGTACGT
2811
glr-2b un... CATCAA
F7-R CATCAA

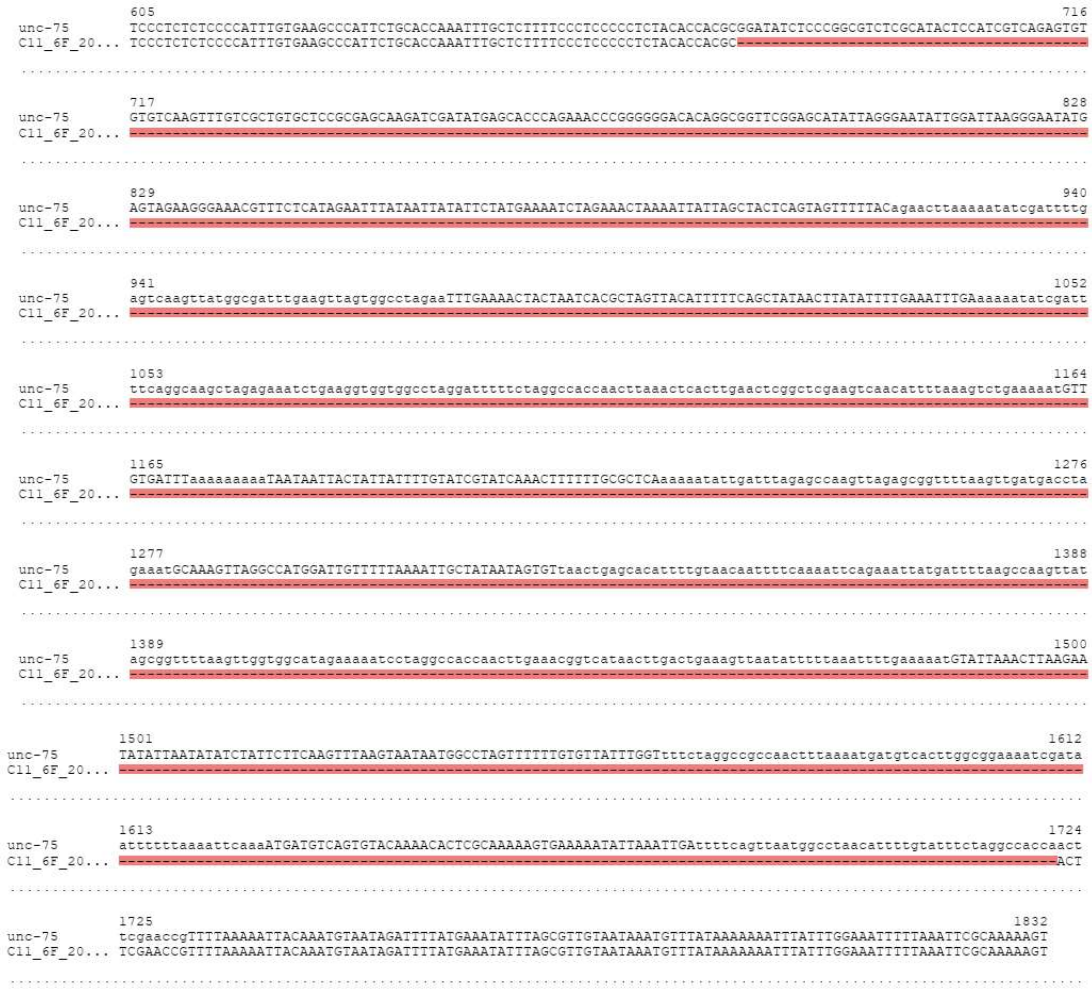
gpa-1(ix265)


2189 2270
gpa-1 ATAGAAAAGTCAAACATTCGAATGATGTTTGATTAGTAATCGCATTTCCATGGTGCATCAATGCTATCAATGCAATCGGCGG
gpa1-F1-9 ATAGAAAAGTCAAACATTCGAATGATGTTTGATTAGTAATCGCATTTCCATGG----------
2271 2352
gpa-1 CAGGTGGTCAGGCAGACACACAAAACAGGGTGGCAGGCTAAACCCGCGAGGTGTCGGCTGTCGTGTACACACTAACCCCGTC
gpa1-F1-9 ----------
2353 2434
gpa-1 AGCACTCTGCTTACATCGAGACCATACTATACAGAATCATTTCTGCAGTATGTATCTCGTAATTCCCATCAAATTGGTAGAG
gpa1-F1-9 ----------
2435 2516
gpa-1 ATGCCATCCGCGATGAGTGGGAGACGGGCGAGCCGCGAGCGTGAAACAGTTCTTCGGAACTTGATGACCGCACGGAGCCGAT
gpa1-F1-9 ----------
2517 2598
gpa-1 GAATTTTGGAGCCGTAGTGGCTTGCTCTTCTGAGCGGATATAATTGAACAAGATTTCAACCTGATGTTTTTTAAACGTTTCT
gpa1-F1-9 ----------
2599 2680
gpa-1 TTAATGCATTTGCGGGTGTTAGTGAACTCAAGTGGGTAGATTTTAATTCCTTCTATTTTGGGTTTCGCTACAGCTGTGGTGC
gpa1-F1-9 ----------
2681 2762
gpa-1 TTTCAACTTTTTGGTTGGGTAACTAGACCTTTCTAGAAACATGTTTTCGTGCCGATACCAGAATGAAATGAGGTTGTCACCA
gpa1-F1-9 ----------ATGAGGTTGTCACCA
2763 2784
gpa-1 CCTATTTGTAAGAAAAATGTTC
gpa1-F1-9 CCTATTTGTAAGAAAAATGTTC

unc-76(ix266)


611 722
unc-75 TCTCCCCATTTGTGAAGCCCATTCTGCACCAAATTTGCTCTTTTCCCTCCCCCTCTACACCACGCGGATATCTCCCGGCGTCTCGCATACTCCATCGTCAGAGTGTGTGTCA
B11_4R_20... TCTCCCCATTTGTGAAGCCCATTCTGCACCAAATTTGCTCTTTTCCCTCCCCCTCTACACCACGT
723 834
unc-75 AGTTTGTCGCTGTGCTCCGCGAGCAAGATCGATATGAGCACCCAGAAACCCGGGGGGACACAGGCGGTTCGGAGCATATTAGGGAATATTGGATTAAGGGAATATGAGTAGA
B11_4R_20...
835 946
unc-75 AGGGAAACGTTTCTCATAGAATTTATAATTATATTCTATGAAAATCTAGAAACTAAAATTATTAGCTACTCAGTAGTTTTTACagaacttaaaaatatcgattttgagtcaa
B11_4R_20...
947 1058
unc-75 gttatggcgatttgaagttagtggcctagaaTTTGAAAACTACTAATCACGCTAGTTACATTTTTCAGCTATAACTTATATTTTGAAATTTGAaaaaatatcgattttcagg
B11_4R_20...
1059 1170
unc-75 caagctagagaaatctgaaggtggtggcctaggatttttctaggccaccaacttaaactcacttgaactcggctcgaagtcaacattttaaagtctgaaaaatGTTGTGATT
B11_4R_20...
1171 1282
unc-75 TaaaaaaaaaTAATAATTACTATTATTTTGTATCGTATCAAACTTTTTTGCGCTCAaaaaaatattgatttagagccaagttagagcggttttaagttgatgacctagaaatG
B11_4R_20...
1283 1394
unc-75 CAAAGTTAGGCCATGGATTGTTTTTAAAATTGCTATAATAGTGTtaactgagcacattttgtaacaattttcaaaattcagaaatttatgattttaagccaagttatagcggt
B11_4R_20...
1395 1506
unc-75 tttaagttggtggcatagaaaaatcctaggccaccaacttgaaacggtcataacttgactgaaagttaatatttttaaatttttgaaaaatGTATTAAACTTAAGAATATATT
B11_4R_20...
1507 1618
unc-75 AATATATCTATTCTTCAAGTTTAAGTAATAATGGCCTAGTTTTTTGTGTTATTTGGTtttctaggccgccaactttaaaatgatgtcacttggcggaaaatcgataatttttt
B11_4R_20... AATGTAATTAC
1619 1730
unc-75 taaaattcaaaaTGATGTCAGTGTACAAAACACTCGCAAAAAGTGAAAAATATTAAATTGAttttcagttaatggcctaacattttgtatttctaggccaccaacttcgaac
B11_4R_20... AAATG TAATAGC ACTTCGAAC
1731 1786
unc-75 cgTTTTAAAAATTACAAATGTAATAGATTTTATGAAATATTTAGCGTTGTAATAAA
B11_4R_20... CGTTTTAAAAATTACAAATGTAATAGATTTTATGAAATATTTAGCGTTGTAATAAA

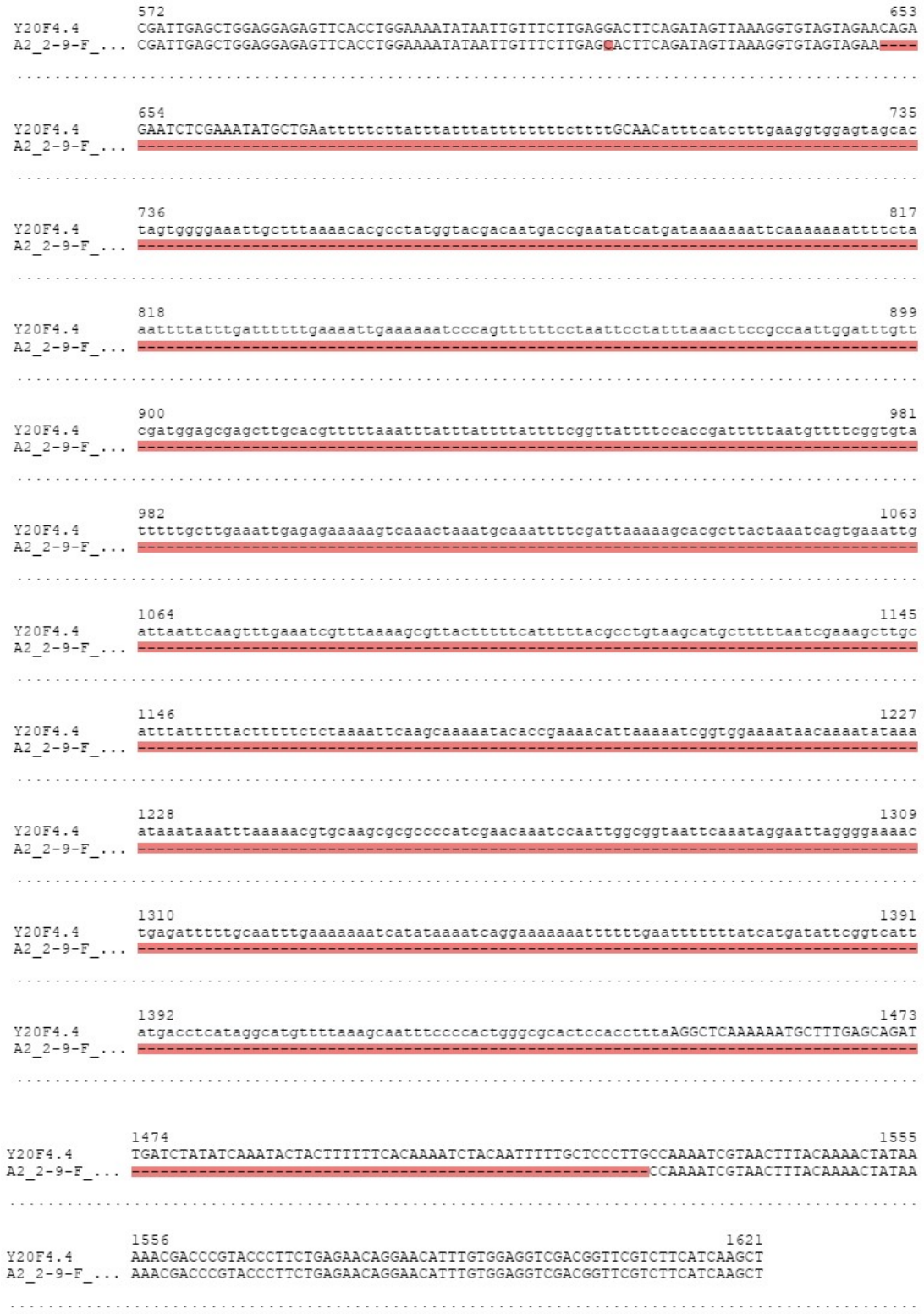
unc-75(ix267)


605 716
unc-75 TCCCTCTCTCCCCATTTGTGAAGCCCATTCTGCACCAAATTTGCTCTTTTCCCTCCCCCTCTACACCACGCGGATATCTCCCGGCGTCTCGCATACTCCATCGTCAGAGTGT
C11_6F_20... TCCCTCTCTCCCCATTTGTGAAGCCCATTCTGCACCAAATTTGCTCTTTTCCCTCCCCCTCTACACCACGC
717 828
unc-75 GTGTCAAGTTTGTCGCTGTGCTCCGCGAGCAAGATCGATATGAGCACCCAGAAACCCGGGGGGACACAGGCGGTTCGGAGCATATTAGGGAATATTGGATTAAGGGAATATG
C11_6F_20...
829 940
unc-75 AGTAGAAGGGAAACGTTTCTCATAGAATTTATAATTATATTCTATGAAAATCTAGAAACTAAAATTATTAGCTACTCAGTAGTTTTTACagaacttaaaaatatcgattttg
C11_6F_20...
941 1052
unc-75 agtcaagttatggcgatttgaagttagtggcctagaaTTTGAAAACTACTAATCACGCTAGTTACATTTTTCAGCTATAACTTATATTTTGAAATTTGAaaaaatatcgatt
C11_6F_20...
1053 1164
unc-75 ttcaggcaagctagagaaatctgaaggtggtggcctaggatttttctaggccaccaacttaaactcacttgaactcggctcgaagtcaacattttaaagtctgaaaaatGTT
C11_6F_20...
1165 1276
unc-75 GTGATTTaaaaaaaaaTAATAATTACTATTATTTTGTATCGTATCAAACTTTTTTGCGCTCAaaaaatattgatttagagccaagttagagcggttttaagttgatgaccta
C11_6F_20...
1277 1388
unc-75 gaaatGCAAAGTTAGGCCATGGATTGTTTTTAAAATTGCTATAATAGTGTtaactgagcacattttgtaacaattttcaaaattcagaaattatgattttaagccaagttat
C11_6F_20...
1389 1500
unc-75 agcggttttaagttggtggcatagaaaaatcctaggccaccaacttgaaacggtcataacttgactgaaagttaatatttttaaattttgaaaaatGTATTAAACTTAAGAA
C11_6F_20...
1501 1612
unc-75 TATATTAATATATCTATTCTTCAAGTTTAAGTAATAATGGCCTAGTTTTTTGTGTTATTTGGTtttctaggccgccaactttaaaatgatgtcacttggcggaaaatcgata
C11_6F_20...
1613 1724
unc-75 atttttaaaattcaaaATGATGTCAGTGTACAAAACACTCGCAAAAAGTGAAAAATATTAAATTGAttttcagttaatggcctaacatttgtatttctaggccaccaact
C11_6F_20... ACT
1725 1832
unc-75 tcgaaccgTTTTAAAAATTACAAATGTAATAGATTTTATGAAATATTTAGCGTTGTAATAAATGTTTATAAAAAAATTTATTTGGAAATTTTTAAATTCGCAAAAAGT
C11_6F_20... TCGAACCGTTTTAAAAATTACAAATGTAATAGATTTTATGAAATATTTAGCGTTGTAATAAATGTTTATAAAAAAATTTATTTGGAAATTTTTAAATTCGCAAAAAGT

iglr-3(ix268)


2939 3050
iglr-3 tgccgaatatccggaaaaacggcaaatcagcaaatttggcggaattgaaaagttccggcaaatcggcagattgccggaactgaaaattttcaggaaaatcggcaaaccggcaac
seq tgccgaatatccggaaaaacggcaaatcagcaaat
3051 3162
iglr-3 ttgccggaattgaaatttttggaaaatcatttctttaaacatttttttttactgtaattttgttttaattattcgtattaaaacattgtaggcgtcgaaacatgcaacattgc
seq
3163 3274
iglr-3 tctggatttcctcaaaagccaatttttattccaaatttCCAAAttttttttgtcagagtgtcccatttcggctcgatctacgtagatctacaaaaaaatgcgggagaaaagac
seq
3275 3386
iglr-3 gcagagtttccaactgatttcgtatgattaacaacgtgctgacgtcacatatttttggtcaaaaaaatcccgcatttttttgtagatcaaaccacgtggtgacagACGCCGTT
seq
3387 3498
iglr-3 TGAATCACCCCCCCCGCCACTGACAAAGTCTCAATTTCATGTTGTTAATGTGTGAAACAGGCACTTGAGAGCTGCCACTCGAGTGGGGTCCCGTGTTTGTTTCCATTCAAACA
seq
3499 3610
iglr-3 TCTGGCCATTAGGCATTAGACCCCCCTCTCCCCCTCTTGGTAGATCACACACACTGACAGTTAATTTCGAATCAAAATGAGAGAGATGAGCCTAgggggggggggAGTGTATT
seq
3611 3722
iglr-3 CAATGAAAAGAGCACAACCACCACTTTAAATTAATCCCATGAGGGTAGAGACTTGGCTACAGAGTTTCACATTCCGTTGCATGAGGATTAGGGCGAAAGCTAATTTTTTGAA
seq
3723 3834
iglr-3 GAAATCCAAATTTTGTTCCCCCGAATCGGCCGAATTCAAATAGGAACTAGAACAACTCGGAGAGCAACTCCCCAGAAATTTCTTGTTATAGAAAATTGTGAGCAGAGATTGG
seq
3835 3946
iglr-3 TAGAATTTTGAAATCTGAccccccccccccccACTTCTCCATTCCATTTTTCCGGTCTCGTATTCTATACTGCTTCGGACCTTTCCCACGGAAAACATATTTCCGAAGATTTT
seq
3947 4058
iglr-3 CACAGAATTACACGACTCAATGCTATGGCAGTAcacggatttctgacttccctcataaattgaaatggaagagtttttgccgaactagaccattttggctcggtcatgtctg
seq
4059 4170
iglr-3 aggtagatttacggcgcgttgcgtgtcgcgtcgcggctcgattatagttgtaaaactaaatgtatttgtccgcgtggagtacacgactttcccagcccggcgggcgattgtc
seq
4171 4282
iglr-3 aatggagcgcgaaaaaattcaatgaggaaggccagaagcccgtgACTTATGCATTTAACGAAAAGGTGGAGCTTTCTCATTGAGGCTaaaatttttagatttttcaagattt
seq TTGAGGCTAAAATTTTTAGATTTTTCAAGATTT
4283
iglr-3 tc
seq TC

Y20F4.4(ix269)



7.2 List of circRNAs showing miRNA sponge functions

Table 7.1 Summary of circRNAs showing miRNA sponge functions.

Name	Tissues or related diseases	miRNA	Target	Ref.
circRNA.2837	Rat sciatic nerve from the rat SNI model	miR-34	-	(*196*)
hsa_circ_0025039	Human malignant melanoma	miR-198	CDK4	(*197*)
hsa_circ_0000993	Human gastric cancer cells	miR-214-5p	-	(*198*)
hsa_circ_0052112	Human breast cancer	miR-125-5p	VEGF-A	(*199*)
circ-VANGL1	Human bladder cancer	miR-605-3p	VANGL1	(*200*)
circRNA_0046367	Human liver	miR-33	ABCA1	(*201*)
circZfp609	mouse myoblast cell line (C2C12)	miR-194-5p	BCLAF1	(*202*)
circRNA-000911	Human breast cancer	miR-499a	NF-B	(*203*)
hsa_circ_0020123	Human non-small cell lung cancer cell	miR-144	ZEB1, EZH2	(*204*)
circRNA_0084043	Human melanoma	miR-153-3p	Snail	(*205*)
circRNA_100284	Arsenite-transformed human hepatic epithelial (L-02) cells	microRNA-217	EZH2	(*206*)
hsa_circ_0032462, hsa_circ_0028173, hsa_circ_0005909	Human osteosarcoma	miR-338-3p; miR-142-5p	CADM1	(*207*)
hsa_circ_100395	Human lung cancer	miR-1228	TCF21	(*208*)
circDLGAP4	Mouse brain	miR-143	HECTD1	(*209*)
circRNA_100269	Human gastric cancer	miR-630	-	(*210*)
circ_0026344	Human colorectal cancer	miR-21, miR-31	-	(*211*)
circNFIX	Human glioma cell	miR-34a-5p	NOTCH1	(*212*)
circCEP128	Human bladder carcinoma	miR-145-5p	SOX11	(*213*)
circDOCK1	Human oral squamous cell carcinoma	miR-196a-5p	BIRC3	(*214*)
circRNA_104670	Human nucleus pulposus (NP) tissues	miR-17-3p	MMP-2	(*215*)
circVMA21	Human nucleus pulposus (NP) tissues	miR-200c	XIAP	(*216*)
circSVIL	Chicken	miR-203	c-JUN, MEF2C	(*217*)
has_circ_0015758 (circ-CFH)	Human glioma	miR-149	AKT1	(*218*)
circFBLIM1	Human hepatocellular cancer tissue	miR-346	FBLIM1	(*219*)
circSMAD2	Human hepatocellular carcinoma	miR-629	-	(*220*)
circ-ITCH	Human bladder cancer Human ovarian cancer	miR-17, miR-224; miR-145	p21, PTEN; RASA1	(*221*), (*222*)
circIRAK3	Human breast cancer	miR-3607	FOXC1	(*223*)
hsa_circ_0001564	Human osteosarcoma	miR-29c-3p	-	(*224*)
circRNA-000284	Human cervical cancer cells	miR-506	Snail-2	(*225*)
circRNA8924	Human cervical cancer cells	miR-518d-5p, miR-519-5p	CBX8	(*226*)
hsa_circ_0000977	Human pancreatic ductal adenocarcinoma	miR-874-3p	PLK1	(*227*)
circRNA_0046366	Human hepatic steatosis	miR-34a	PPAR alpha	(*228*)
circEPSTI1 (hsa_circRNA_000479)	Human triple-negative breast cancer	miR-4753, miR-6809	BCL11A	(*229*)
circGFRA1	Human triple-negative breast cancer	miR-34a	GFRA1	(*230*)
circRNA_010567	Mice myocardial fibrosis	miR-141	TGF-beta 1	(*231*)

circBIRC6	undifferentiated human embryonic stem cells (hESCs)	miR-34a, miR-145	multiple	(*83*)
circRNA_000203	Human cardiac fibroblasts	miR-26b-5p	Col1a2, CTGF	(*232*)
circZNF609	Hirschsprung's disease	miR-150-5p	AKT3	(*233*)
circRNA-100338	Human hepatocellular carcinoma	miR-141-3p		(*234*)
circMFACR	Mice cardiomyocyte	miR-652-3p	MTP18	(*235*)
circMTO1	Human hepatocellular carcinoma	miR-9	p21	(*236*)
circMTO1	Glioblastoma	miR-630	-	(*237*)
circMTO1	Bladder cancer	miR-221	E-cadherin/N-cadherin	(*238*)
circRNA_100290	Human oral squamous cell carcinomas	miR-29	CDK6	(*239*)
circTCF25	Human bladder carcinoma	miR-103a-3p; miR-107	CDK6	(*240*)
hsa_circ_001569	Human colorectal cancer	miR-145	E2F5, BAG4 , FMNL2	(*241*)
circHIPK3	Human tissues	miR-124	IL6R, DLX2	(*82*)
has_circ_101280	Human hepatocellular carcinoma cells	miR-375	JAK2	(*242*)
BCRC3(circPSMD1)	Human bladder cancer	miR-182-5p	p27	(*243*)
circASAMTS14	Human hepatocellular carcinoma	miR-572	RCAN1	(*244*)
circYAP1	Human gastric cancer	miR-367-5p	p27(Kip1)	(*245*)
circUBXN7	Bladder cancer	miR-1247-3p	B4GALT3	(*246*)
circ104075	Human hepatocellular carcinoma	miR-582-3p	YAP	(*247*)
circUVRAG	Bladder Cancer	miR-233	FGFR2	(*248*)
circ_0067934	Cervical cancer	miR-545	EIF3C	(*249*)
circPAN3	acute myeloid leukemia (AML) cell line	miR-153-5p/miR-183-5p	X-linked inhibitor of apoptosis protein (XIAP)	(*250*)
circFGFR2	Chicken embryo skeletal muscle development	miR-133a-5p; miR-29b-1-5p	-	(*251*)
circLARP4	Gastric cancer	miR-424-5p	large tumor suppressor kinase 1 (LATS1)	(*252*)
circABCB10	Non-small cell lung cancer cell	miR-1252	Forkhead box 2 (FOXR2)	(*253*)
circFAT1(e2)	Gastric cancer	miR-548g	RUNX1	(*95*)
hsa_circ_0074362	Glioma	miR-1236-3p	HOXB7	(*254*)
circDYM	Major depressive disorder patient and CUS & LPS mouse model	miR-9	HECTD1	(*255*)
circFNDC3B	Human bladder cancer	miR-1178-3p	G3BP2	(*256*)
circANKS1B	Breast cancer	miR-148a-3p; miR-152-3p	USF1	(*257*)
has_circ_0078710 (circTHBS2)	Human hepatocellular carcinoma	miR-31	-	(*258*)
circZNF609	Renal carcinoma	miR-138-5p	FOXP4	(*259*)
circZNF609	Retinal neurodegeneration	miR-615	METRN	(*260*)
hsa_circ_0000488(circRNA-DLEU2)	Human acute myeloid leukemia cell	miR-496	PRKACB	(*261*)
circNT5E	Glioblastoma	miR-422a	-	(*262*)
circMMP9	Glioblastoma multiforme	miR-124	-	(*263*)

circPCNXL2	Clear cell renal cell carcinoma	miR-153	ZEB2	(*264*)
hsa_circ_0136666	Colorectal cance	miR-136	SH2B1	(*265*)
hsa_circ_0076248	Glioma	miR-181a	SIRT1	(*266*)
circFAT1	Human osteosarcoma	miR-375	Yes-associated protein 1 (YAP1)	(*267*)
circUVRAG	Bladder cancer	miR-223	FGFR2	(*268*)
hsa_circ_0004015 (circCDK14)	Non-small cell lung cancer	miR-1183	PDPK1	(*269*)
circ_0034642	Glioma	miR-1205	BATF3	(*270*)
Hsa_circ_0060060, circEIF6	Human thyroid carcinoma	miR-144-3p	TGF-α	(*271*)
circPVT1	Non-small cell lung cancer	miR-125b	E2F2	(*272*)
circPVT1	Non-small cell lung cancer	miR-497	Bcl-2	(*273*)
circNF1	Gastric carcinoma	miR-16	MAP7, AKT3	(*274*)
circ_Lrp6	Vascular smooth muscle cell	miR-145	ITGβ8, FASCIN, KLF4, Yes1, Lox	(*275*)
circZKSCAN1	Bladder cancer	miR-1178-3p	p21	(*276*)
circHIPK3	Chiken myoblast cells	miR-30a-3p	MEF2C	(*277*)
circTP63	Lung squamous cell carcinoma	miR-873-3p	FOXM1, CENPA, CENPB	(*278*)
circFOXP1	Mesenchymal stem cell	miR-17-3p; miR-127-5p	WNT5A; WNT3A	(*279*)
hsa_circ_0001073 (circACVR2A)	Bladder cancer cell	miR-626	EYA4	(*280*)
hsa_circ_0000199 (circAKT3)	Gastric cancer	miR-198	PIK3R1	(*281*)
circRNA CBL.11	Colorectal cancer	miR-6778-5p	p53	(*282*)
cPWWP2A	Vascular pericytes and endothelial cells	miR-579	angiopoietin 1, occludin, SIRT1	(*283*)
circRAPGEF5	Papillary thyroid cancer	miR-198	FGFR1	(*284*)
circSL8A1	Bladder cancer	miR-130b; miR-494	PTEN	(*285*)
cTFRC	Bladder cancer	miR-107	TFRC	(*286*)
circTADA2A	Osteosarcoma	miR-203s-3p	CREB3	(*287*)
circTADA2A-E6	Triple-negative breast cancer	miR203a-3p	SOCS3	(*288*)
circAnks1a	Rat spinal cord	miR-324-3p	VEGFB	(*289*)
circNRIP1	Gastric cancer	miR-149-5p	AKT1/mTOR	(*290*)
circNHSL1	Gastric cancer	miR-1306-3p	SIX1/vimentin	(*291*)
circ-MALAT1	Hepatocellular cancer stem cell	miR-6887-3p	JAK2	(*64*)
circ-ERBIN	Colorectal cancer	miR-125a-5p; miR-138-5p	4EBP1	(*292*)
cRAPGEF5	Renal cell carcinoma	miR-27a-3	TXNIP	(*293*)
circSnx5	Dendritic cell	miR-544	SOCS1	(*294*)
circ-ADAM9	Endothelial progenitor cells	miR-20a-5p	PTEN; ATG7	(*295*)
circSEMA5A	Bladder cancer	miR-330-5p	ENO1	(*296*)
circEYA1	Cervical adenocarcinoma	miR-528-3p	CXCL14	(*297*)
circNR3C2	Triple-negative breast cancer	miR-513a-3p	HRD1	(*298*)
circNEIL3	Pancreatic ductal adenocarcinoma	miR-432-5p	ADAR1	(*299*)
circST6GALNAC6	Bladder cancer	miR-200a-3p	STMN1	(*300*)
circPGR	Estrogen receptor-positive breast cancer	miR-301-5p	-	(*301*)
circPARP4	Glioblastoma	miR-125a-5p	FUT4	(*302*)

7.3 Code scripts

7.3.1 Sequence alignment by STAR

Joint map, mate1 map and mate2 map by STAR, using whole_3 data as an example

```
###DCC_jointMap

#!/bin/bash

#SBATCH --job-name=whole_3_joint_map

#SBATCH --partition=compute

#SBATCH --time=2:00:00

#SBATCH --mem-per-cpu=4G

#SBATCH --ntasks=1

#SBATCH --cpus-per-task=12

#SBATCH --mail-user=dong.cao@oist.jp

#SBATCH --mail-type=FAIL,END


"/work/MaruyamaU/dongcao/apps/STAR/bin/Linux_x86_64/STAR" --runThreadN 24 \

     --genomeDir /work/MaruyamaU/dongcao/RNAseq/starindex/starindex_cel235/ \

     --outSAMtype BAM SortedByCoordinate \

     --readFilesIn "/work/MaruyamaU/dongcao/RNAseq/1-
1_MaruyamaU_ID361/ce_whole_3_R1.fastq.gz" "/work/MaruyamaU/dongcao/RNAseq/1-
1_MaruyamaU_ID361/ce_whole_3_R2.fastq.gz" \

     --readFilesCommand zcat \

     --outFileNamePrefix whole_3_ \

     --outReadsUnmapped Fastx \

     --outSJfilterOverhangMin 15 15 15 15 \

     --alignSJoverhangMin 15 \

     --alignSJDBoverhangMin 15 \

     --outFilterMultimapNmax 20 \

     --outFilterScoreMin 1 \

     --outFilterMatchNmin 1 \

     --outFilterMismatchNmax 2 \

     --chimSegmentMin 15 \

     --chimScoreMin 15 \

     --chimScoreSeparation 10 \

     --chimJunctionOverhangMin 15 \


###DCC_mate1
```

```
#!/bin/bash
#SBATCH --job-name=whole_3_mate1
#SBATCH --partition=compute
#SBATCH --time=2:00:00
#SBATCH --mem-per-cpu=4G
#SBATCH --ntasks=1
#SBATCH --cpus-per-task=12
#SBATCH --mail-user=dong.cao@oist.jp
#SBATCH --mail-type=FAIL,END

"/work/MaruyamaU/dongcao/apps/STAR/bin/Linux_x86_64/STAR" --runThreadN 24 \
     --genomeDir /work/MaruyamaU/dongcao/RNAseq/starindex/starindex_cel235/ \
     --outSAMtype None \
     --readFilesIn "/work/MaruyamaU/dongcao/RNAseq/1-
1_MaruyamaU_ID361/ce_whole_3_R1.fastq.gz" \
     --readFilesCommand zcat \
     --outFileNamePrefix whole_3_mate1_ \
     --outReadsUnmapped Fastx \
     --outSJfilterOverhangMin 15 15 15 15 \
     --alignSJoverhangMin 15 \
     --alignSJDBoverhangMin 15 \
     --seedSearchStartLmax 30 \
     --outFilterMultimapNmax 20 \
     --outFilterScoreMin 1 \
     --outFilterMatchNmin 1 \
     --outFilterMismatchNmax 2 \
     --chimSegmentMin 15 \
     --chimScoreMin 15 \
     --chimScoreSeparation 10 \
     --chimJunctionOverhangMin 15 \

###DCC_mate2
#!/bin/bash
#SBATCH --job-name=whole_3_mate2
#SBATCH --partition=compute
#SBATCH --time=2:00:00
```

```
#SBATCH --mem-per-cpu=4G
#SBATCH --ntasks=1
#SBATCH --cpus-per-task=12
#SBATCH --mail-user=dong.cao@oist.jp
#SBATCH --mail-type=FAIL,END

"/work/MaruyamaU/dongcao/apps/STAR/bin/Linux_x86_64/STAR" --runThreadN 24 \
     --genomeDir /work/MaruyamaU/dongcao/RNAseq/starindex/starindex_cel235/ \
     --outSAMtype None \
     --readFilesIn "/work/MaruyamaU/dongcao/RNAseq/1-
1_MaruyamaU_ID361/ce_whole_3_R2.fastq.gz" \
     --readFilesCommand zcat \
     --outFileNamePrefix whole_3_mate2_ \
     --outReadsUnmapped Fastx \
     --outSJfilterOverhangMin 15 15 15 15 \
     --alignSJoverhangMin 15 \
     --alignSJDBoverhangMin 15 \
     --seedSearchStartLmax 30 \
     --outFilterMultimapNmax 20 \
     --outFilterScoreMin 1 \
     --outFilterMatchNmin 1 \
     --outFilterMismatchNmax 2 \
     --chimSegmentMin 15 \
     --chimScoreMin 15 \
     --chimScoreSeparation 10 \
     --chimJunctionOverhangMin 15 \
```

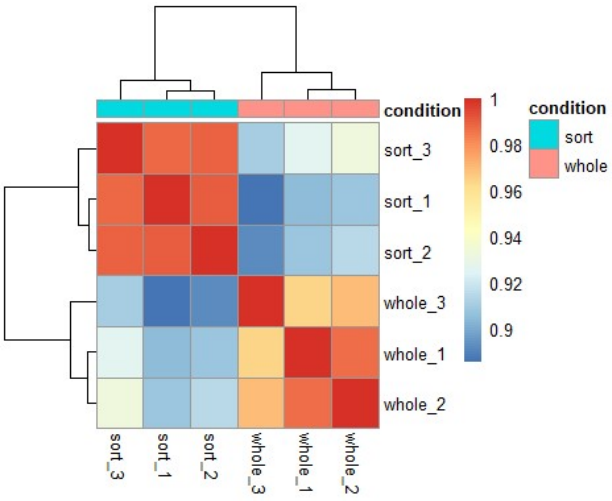
7.3.2 circRNA annotation by DCC

```
### DCC_annotation

#!/bin/bash

#SBATCH --job-name=DCC_test

#SBATCH --partition=compute

#SBATCH --time=10:00:00

#SBATCH --mem-per-cpu=4G

#SBATCH --ntasks=1

#SBATCH --cpus-per-task=12

#SBATCH --mail-user=dong.cao@oist.jp

#SBATCH --mail-type=FAIL,END


module load python/2.7.10

DCC @samplesheet -mt1 @mate1 -mt2 @mate2 -T 24 -O ./testforNroption -D -R
ce11_repeat_file.gtf -an /work/MaruyamaU/dongcao/WBcel235/Annotation/Archives/archive-2015-
07-17-14-28-46/Genes/genes.gtf -Pi -F -M -Nr 1 1 -fg -G -A
/work/MaruyamaU/dongcao/WBcel235/Sequence/WholeGenomeFasta/genome.fa
```

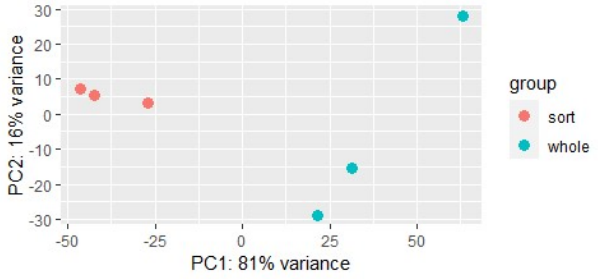
7.3.3 Differential expression analysis of mRNAs using DESeq2

```
library(ggplot2)
library(dplyr)

library(DESeq2)

library(readr)
library(pheatmap)
library(tibble)
library(ggrepel)
library(ggpubr)

#setwd(~/)
setwd("~/RNA-seq analysis")

genecounts_with_names <- read_csv("genecounts with names.csv")

## Parsed with column specification:
## cols(
##   X1 = col_character(),
##   gene = col_character(),
##   sort_1 = col_double(),
##   sort_2 = col_double(),
##   sort_3 = col_double(),
##   whole_1 = col_double(),
##   whole_2 = col_double(),
##   whole_3 = col_double(),
##   gene_name = col_character(),
##   sequence_name = col_character(),
##   sort_sum = col_double(),
##   whole_sum = col_double()
## )

read_counts <- data.frame(genecounts_with_names[, 3:8], row.names =
genecounts_with_names$gene)

condition <- c("sort","sort","sort", "whole", "whole", "whole")
batch <- c("b1","b2","b3","b1","b2","b3")
NW1229_metadata <- data.frame(condition, batch)
rownames(NW1229_metadata)<- c("sort_1","sort_2","sort_3","whole_1","whole_2","whole_3")

# check if rownames of metadata are consistant with colnames of readcounts table
all(rownames(NW1229_metadata) == colnames(read_counts))

## [1] TRUE

# Creating DESeq2 object

dds<- DESeqDataSetFromMatrix(countData = read_counts,
                             colData = NW1229_metadata,
                             design = ~ condition)

## converting counts to integer mode

# Counts normalization

dds <- estimateSizeFactors(dds)
sizeFactors(dds)
```

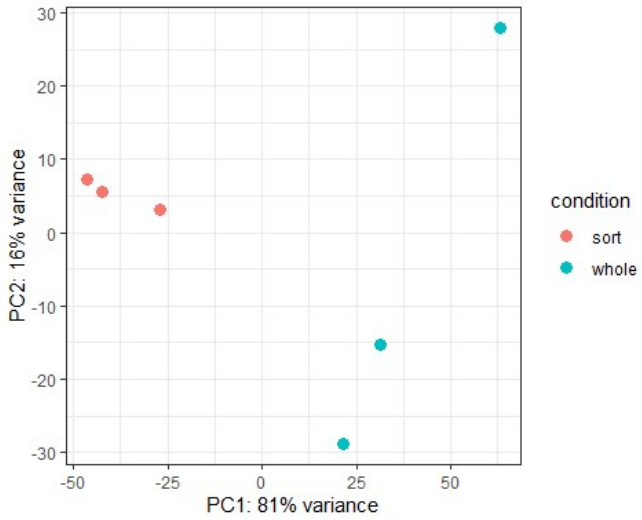
```
##    sort_1    sort_2    sort_3   whole_1   whole_2   whole_3
## 1.1241306 1.1509995 1.4451648 0.7039952 0.9837595 0.7715272

normalized_read_counts <-counts(dds, normalized = T)

# log transformation: variance stabilizing transformation(VST)

vsd <- vst(dds, blind = TRUE)

# Extract the vst matrix from the object
vsd_matrix <- assay(vsd)
# Compute pairwise correlation values
vsd_cor <- cor(vsd_matrix)
vsd_cor

##          sort_1    sort_2    sort_3   whole_1   whole_2   whole_3
## sort_1  1.0000000 0.9897823 0.9880901 0.9038740 0.9076862 0.8855246
## sort_2  0.9897823 1.0000000 0.9893364 0.9084163 0.9149492 0.8913440
## sort_3  0.9880901 0.9893364 1.0000000 0.9259891 0.9324742 0.9102163
## whole_1 0.9038740 0.9084163 0.9259891 1.0000000 0.9864594 0.9636024
## whole_2 0.9076862 0.9149492 0.9324742 0.9864594 1.0000000 0.9701827
## whole_3 0.8855246 0.8913440 0.9102163 0.9636024 0.9701827 1.0000000

# Plot heatmap
pheatmap(vsd_cor, annotation = select(NW1229_metadata, condition))
```



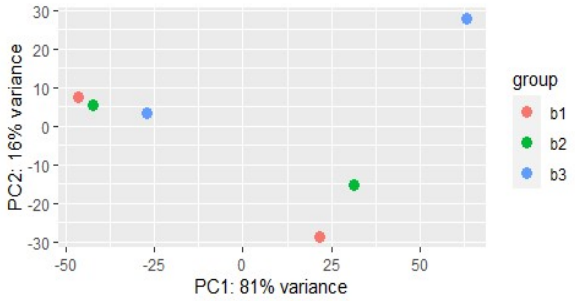

```
# Plot PCA
plotPCA(vsd, intgroup="condition")
```



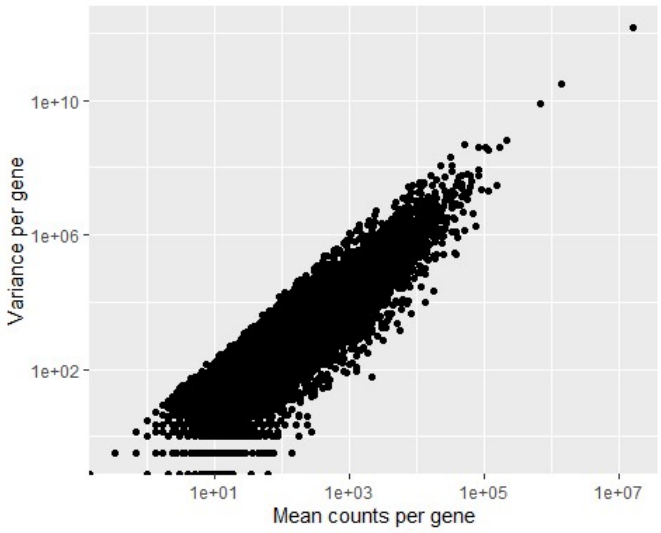

```
##modify PCA plot using ggplot
pcaData <- plotPCA(vsd, intgroup="condition", returnData=TRUE)
percentVar <- round(100 * attr(pcaData, "percentVar"))

ggplot(pcaData, aes(PC1, PC2, color=condition)) +
  geom_point(size=3) +
  xlab(paste0("PC1: ",percentVar[1],"% variance")) +
  ylab(paste0("PC2: ",percentVar[2],"% variance")) +
  theme_bw()
```



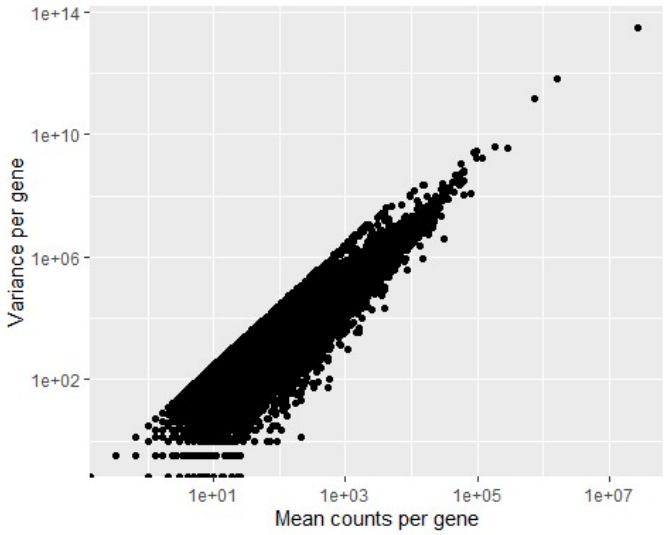

```
plotPCA(vsd, intgroup="batch")
```



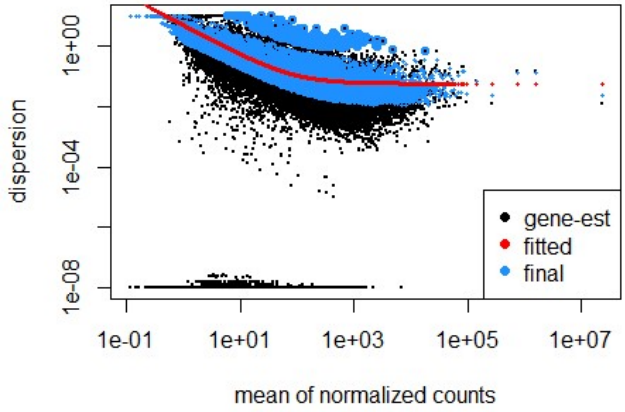

```
## Run analysis

dds <- DESeq(dds)

## using pre-existing size factors

## estimating dispersions

## gene-wise dispersion estimates

## mean-dispersion relationship

## final dispersion estimates

## fitting model and testing

## mean-variance relationship

# Syntax for apply():  apply(data, rows/columns, function_to_apply)
# Calculating mean for each gene (each row) in sort group
mean_counts_sort <- apply(read_counts[, 1:3], 1, mean)
# Calculating variance for each gene (each row)
variance_counts_sort <- apply(read_counts[, 1:3], 1, var)

# Plotting relationship between mean and variance of sort group

df_sort <- data.frame(mean_counts_sort, variance_counts_sort)
ggplot(df_sort) +
geom_point(aes(x=mean_counts_sort, y=variance_counts_sort)) +
scale_y_log10() +
scale_x_log10() +
xlab("Mean counts per gene") +
ylab("Variance per gene")
```



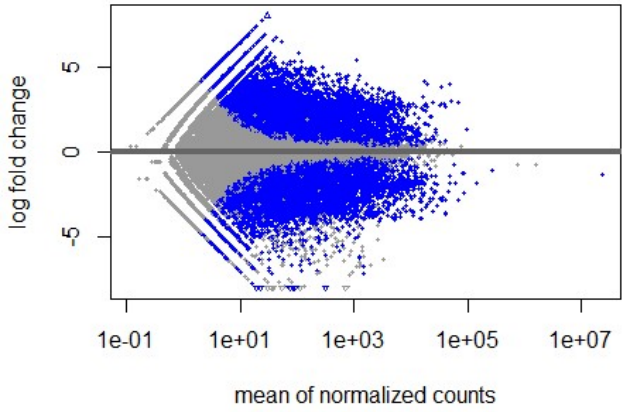

```
# Calculating mean for each gene (each row) in whole group
mean_counts_whole <- apply(read_counts[, 4:6], 1, mean)
# Calculating variance for each gene (each row)
variance_counts_whole <- apply(read_counts[, 4:6], 1, var)

# Plotting relationship between mean and variance of sort group

df_whole <- data.frame(mean_counts_whole, variance_counts_whole)
ggplot(df_whole) +
geom_point(aes(x=mean_counts_whole, y=variance_counts_whole)) +
scale_y_log10() +
scale_x_log10() +
xlab("Mean counts per gene") +
ylab("Variance per gene")
```



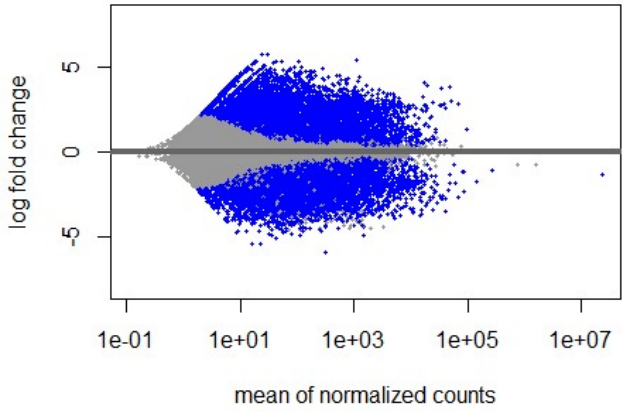

```
## Plot dispersion estimates
plotDispEsts(dds)
```




```
# Results extraction
result <- results(dds,
                  contrast = c("condition", "sort", "whole"),
                  alpha = 0.05)

plotMA(result, ylim = c(-8,8))
```

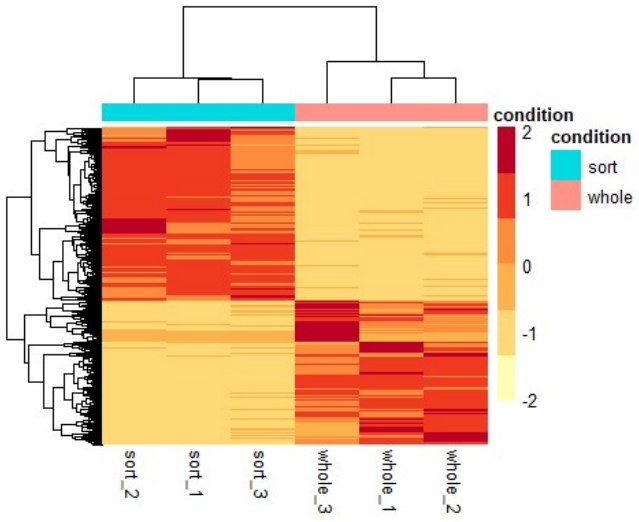



```
### LFC shrinkage with "ashr" (Stephens, M. (2016) False discovery rates: a new deal.
Biostatistics, 18:2.    https://doi.org/10.1093/biostatistics/kxw041)
result <- lfcShrink(dds,
                    type = "ashr",
                    contrast = c("condition", "sort", "whole"),
                    res = result)

## using 'ashr' for LFC shrinkage. If used in published research, please cite:
##     Stephens, M. (2016) False discovery rates: a new deal. Biostatistics, 18:2.
##     https://doi.org/10.1093/biostatistics/kxw041

plotMA(result, ylim = c(-8,8))
```

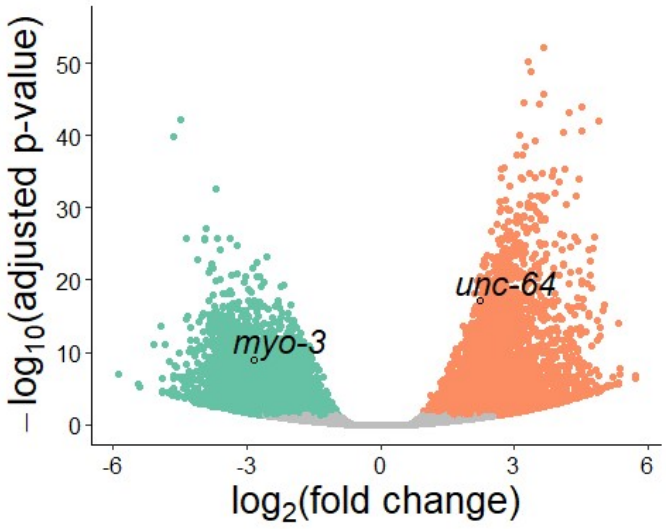



```
# check result table
mcols(result)

## DataFrame with 5 rows and 2 columns
##                       type                                  description
##                <character>                                  <character>
## baseMean       intermediate       mean of normalized counts for all samples
## log2FoldChange      results log2 fold change (MMSE): condition sort vs whole
## lfcSE               results           posterior SD: condition sort vs whole
## pvalue              results       Wald test p-value: condition sort vs whole
## padj                results                         BH adjusted p-values

summary(result)

##
## out of 33552 with nonzero total read count
## adjusted p-value < 0.05
## LFC > 0 (up)       : 7550, 23%
## LFC < 0 (down)     : 6646, 20%
## outliers [1]       : 214, 0.64%
## low counts [2]     : 7644, 23%
## (mean count < 1)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results

### set fold-change threshold(0.585, 1.5-fold)

result <- results(dds,
                  contrast = c("condition", "sort", "whole"),
                  alpha = 0.05,
                  lfcThreshold = 0.585)
result <- lfcShrink(dds,
                    type = 'ashr',
                    contrast = c("condition", "sort", "whole"),
                    res = result)

## using 'ashr' for LFC shrinkage. If used in published research, please cite:
##     Stephens, M. (2016) False discovery rates: a new deal. Biostatistics, 18:2.
##     https://doi.org/10.1093/biostatistics/kxw041

# check result table after setting fold-change threshold
mcols(result)
```

```
## DataFrame with 5 rows and 2 columns
##                       type                                       description
##                <character>                                       <character>
## baseMean       intermediate       mean of normalized counts for all samples
## log2FoldChange      results log2 fold change (MMSE): condition sort vs whole
## lfcSE               results           posterior SD: condition sort vs whole
## pvalue              results       Wald test p-value: condition sort vs whole
## padj                results                         BH adjusted p-values

summary(result)

##
## out of 33552 with nonzero total read count
## adjusted p-value < 0.05
## LFC > 0 (up)       : 5023, 15%
## LFC < 0 (down)     : 4178, 12%
## outliers [1]       : 214, 0.64%
## low counts [2]     : 8918, 27%
## (mean count < 2)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results

### Results extraction

res_all <- data.frame(result) %>% rownames_to_column(var = "WBgene")
res_all$gene_name <- genecounts_with_names$gene_name

res_sig <- subset(res_all, padj < 0.05)

## Subset normalized counts to significant genes
sig_norm_counts <- normalized_read_counts[res_sig$WBgene, ]
## Choose a color palette from RColorBrewer
library(RColorBrewer)
heat_colors <- brewer.pal(6, "YlOrRd")

##Plot heatmap using pheatmap
pheatmap(sig_norm_counts,
         color = heat_colors,
         cluster_rows = T,
         show_rownames = F,
         annotation = select(NW1229_metadata, condition),
         scale = "row")
```



```
### Volcano plot

#Obtain logical vector regarding whether padj values are less than 0.05
res_all <- mutate(res_all, threshold = padj<0.05) %>% mutate(updown = ifelse(threshold == F, 0, ifelse(log2FoldChange > 0, 1,2)))


ggplot(res_all) +
  geom_point(aes(x = log2FoldChange, y = -log10(padj), color = as.factor(updown))) +
  xlab("log2 fold change") +
  ylab("-log10 adjusted p-value") +
  scale_color_manual(values = c("grey","#FC8D62", "#66C2A5")) +
  theme(legend.position = "none",
        plot.title = element_text(size = rel(1.5), hjust = 0.5),
        axis.title = element_text(size = rel(1.25)))

##adjust ylim
ggplot(res_all) +
  geom_point(aes(x = log2FoldChange, y = -log10(padj), color = as.factor(updown))) +
  xlab("log2 fold change") +
  ylab("-log10 adjusted p-value") +
  ylim(c(0,60)) +
  scale_color_manual(values = c("grey","#FC8D62", "#66C2A5")) +
  theme(legend.position = "none",
        plot.title = element_text(size = rel(1.5), hjust = 0.5),
        axis.title = element_text(size = rel(1.25)))

## highlight some genes
library(ggrepel)
gene <- c("unc-64", "myo-3", "glr-2", "gpa-1", "iglr-3","unc-75", "zip-2", "arl-13", "Y20F4.4", "crh-1")
gene2<-c("unc-64", "myo-3")
RBP<- c("C25A1.4", "exc-7", "Y27G11C.9", "mbl-1", "msi-1", "tdp-1", "unc-75", "asd-1", "tiar-3","fox-1", "mec-8", "hrpf-1", "fust-1")

##choose which gene list to be labeled(gene_name %in% ?)
subdata = filter(res_all, gene_name  %in% gene2)
p<- ggplot(res_all, aes(x = log2FoldChange, y = -log10(padj)), label = gene_name) +
  geom_point(aes(color = as.factor(updown))) +
  geom_point(data = subdata, aes(x = log2FoldChange, y = -log10(padj)),shape = 1, color = "black") +
  geom_text_repel(data = subdata, size = 6, label = subdata$gene_name, fontface = "italic") +
  scale_color_manual(values = c("grey","#FC8D62", "#66C2A5")) +
  xlab(expression(log["2"]*"(fold change)")) +
  ylab(expression(-log["10"]*"(adjusted p-value)")) +
  scale_y_continuous(limits = c(0,55), breaks = seq(0,50,10)) +
  theme_pubr()+
  theme(legend.position = "none",
        plot.title = element_text(size = rel(1.5), hjust = 0.5),
        axis.title = element_text(size = rel(1.7)))

p
```

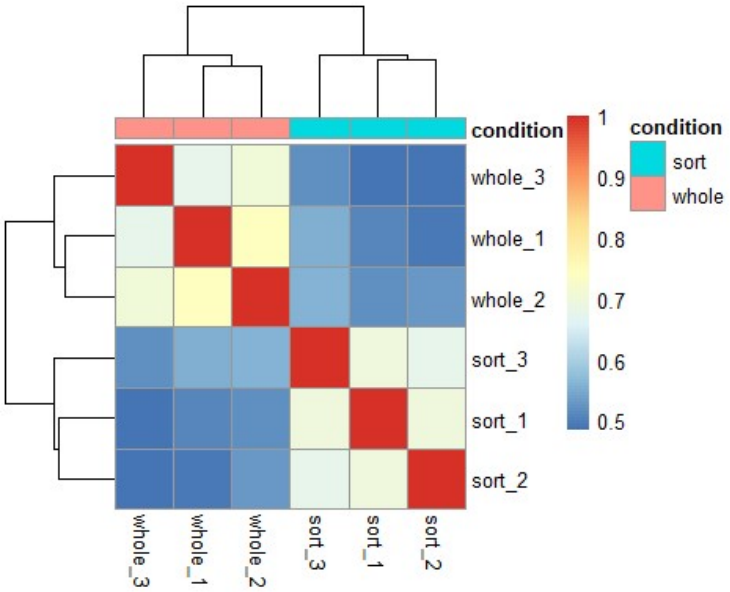



```
##results export

write.csv(res_all,"./linearDE.csv", quote = F, row.names = F)

write.csv(res_all,"./linearDE.txt", quote = F, row.names = F)
```

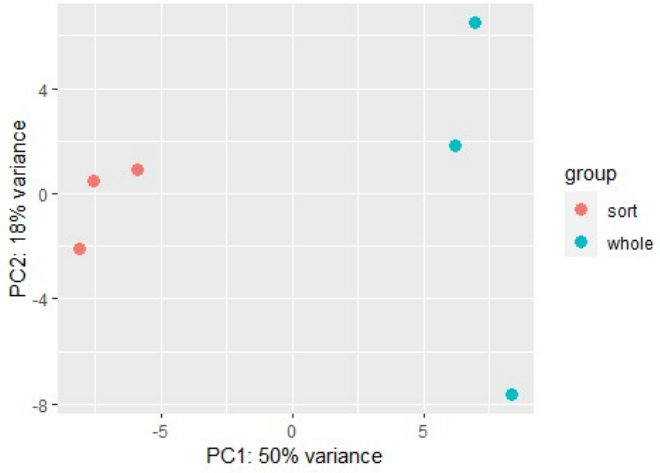
7.3.4 circRNA differential expression analysis using DESeq2

```
library(ggplot2)
library(tidyr)
library(dplyr)
library(DESeq2)
library(readr)
library(pheatmap)
library(tibble)
library(ggrepel)
library(ggpubr)


#setwd(~/)
setwd("~/RNA-seq analysis/circDE/")

circCount <- read_csv("./DCC_output.csv")

## Parsed with column specification:
## cols(
##   circ_ID = col_character(),
##   Chr = col_character(),
##   Start = col_double(),
##   End = col_double(),
##   Gene = col_character(),
##   sort_1 = col_double(),
##   sort_2 = col_double(),
##   sort_3 = col_double(),
##   whole_1 = col_double(),
##   whole_2 = col_double(),
##   whole_3 = col_double()
## )

circCount$sort_sum <- circCount$sort_1 + circCount$sort_2 + circCount$sort_3
circCount$whole_sum <- circCount$whole_1 + circCount$whole_2 + circCount$whole_3

##minimum 3 reads in either group
circCount_3 <- filter(circCount, sort_sum >=3 | whole_sum >=3)
circReads_3 <- data.frame(circCount_3[, 6:11], row.names = circCount_3$circ_ID)


condition <- c("sort","sort","sort", "whole", "whole", "whole")
batch <- c("b1","b2","b3","b1","b2","b3")
metadata <- data.frame(condition, batch)
rownames(metadata)<- c("sort_1","sort_2","sort_3","whole_1","whole_2","whole_3")

# check if rownames of metadata are consistant with colnames of readcounts table
all(rownames(metadata) == colnames(circReads_3))

## [1] TRUE

# Creating DESeq2 object

dds_3<- DESeqDataSetFromMatrix(countData = circReads_3,
                               colData = metadata,
                               design = ~ condition)

## converting counts to integer mode

## Warning in DESeqDataSet(se, design = design, ignoreRank): some variables in
## design formula are characters, converting to factors
```

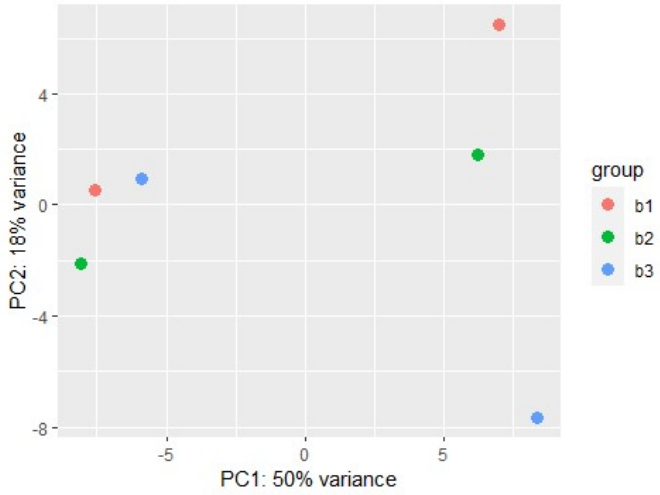
```
# Counts normalization

dds_3 <- estimateSizeFactors(dds_3)
sizeFactors(dds_3)

##    sort_1    sort_2    sort_3   whole_1   whole_2   whole_3
## 1.1157003 1.0815636 1.2950062 0.8437458 1.2698120 0.7202728

normalized_circCounts <-counts(dds_3, normalized = T)

# log transformation: variance stabilizing transformation(VST)

vsd_3 <- varianceStabilizingTransformation(dds_3, blind = T)

# Extract the vst matrix from the object
vsd_matrix_3 <- assay(vsd_3)
# Compute pairwise correlation values
vsd_cor_3 <- cor(vsd_matrix_3)
vsd_cor_3

##          sort_1    sort_2    sort_3   whole_1   whole_2   whole_3
## sort_1  1.0000000 0.6990687 0.6988317 0.5099201 0.5168932 0.4850869
## sort_2  0.6990687 1.0000000 0.6778302 0.4944525 0.5287449 0.4849661
## sort_3  0.6988317 0.6778302 1.0000000 0.5529481 0.5577008 0.5178015
## whole_1 0.5099201 0.4944525 0.5529481 1.0000000 0.7393289 0.6801907
## whole_2 0.5168932 0.5287449 0.5577008 0.7393289 1.0000000 0.7037190
## whole_3 0.4850869 0.4849661 0.5178015 0.6801907 0.7037190 1.0000000

# Plot heatmap
pheatmap(vsd_cor_3, annotation = select(metadata, condition))
```




```
# Plot PCA
plotPCA(vsd_3, intgroup="condition")
```

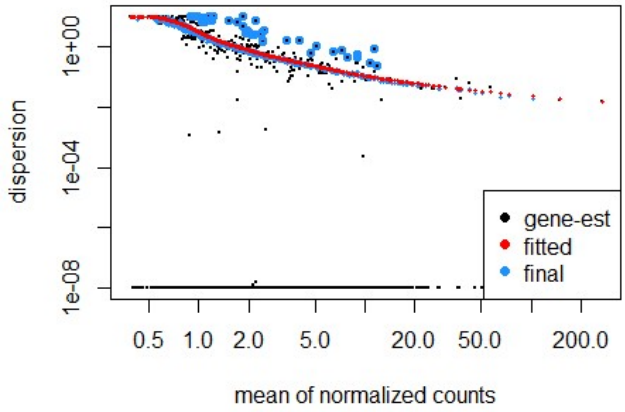



```
plotPCA(vsd_3, intgroup="batch")
```

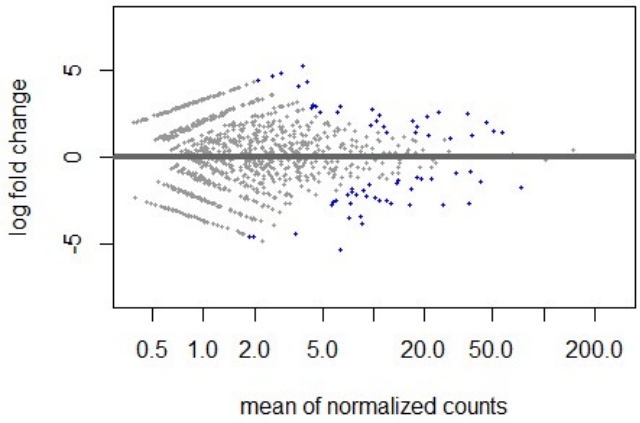



```
## Run analysis

dds_3 <- DESeq(dds_3)

## using pre-existing size factors

## estimating dispersions

## gene-wise dispersion estimates

## mean-dispersion relationship

## -- note: fitType='parametric', but the dispersion trend was not well captured by the
##    function: y = a/x + b, and a local regression fit was automatically substituted.
##    specify fitType='local' or 'mean' to avoid this message next time.

## final dispersion estimates

## fitting model and testing
```

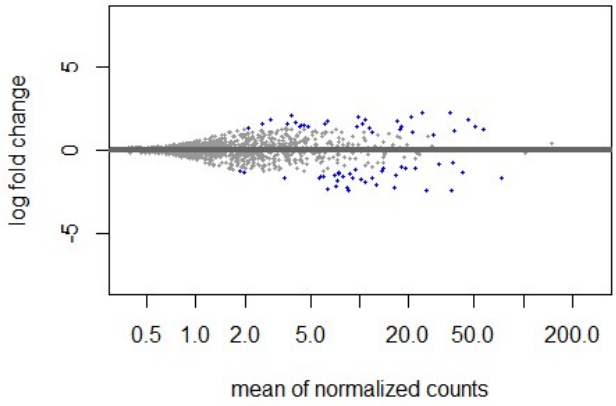
```
## Plot dispersion estimates
plotDispEsts(dds_3)
```




```
# Results extraction
result <- results(dds_3,
                  contrast = c("condition", "sort", "whole"),
                  alpha = 0.05)

plotMA(result, ylim = c(-8,8))
```

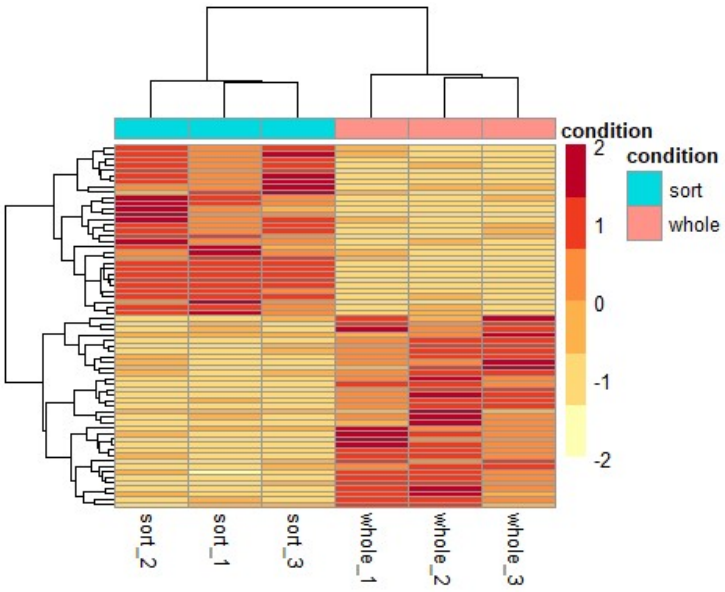



```
### LFC shrinkage
result <- lfcShrink(dds_3,
                    type = "ashr",
                    contrast = c("condition", "sort", "whole"),
                    res = result)

## using 'ashr' for LFC shrinkage. If used in published research, please cite:
##     Stephens, M. (2016) False discovery rates: a new deal. Biostatistics, 18:2.
##     https://doi.org/10.1093/biostatistics/kxw041

plotMA(result, ylim = c(-8,8))
```

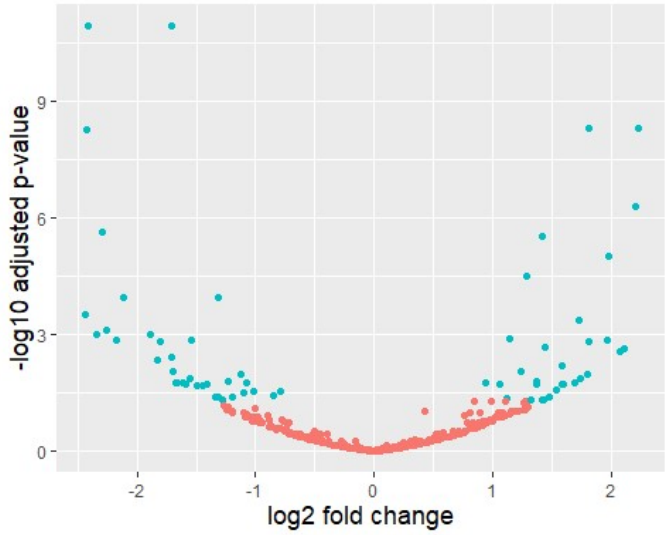



```
# check result table
mcols(result)

## DataFrame with 5 rows and 2 columns
##                       type                                  description
##                <character>                                  <character>
## baseMean       intermediate       mean of normalized counts for all samples
## log2FoldChange      results log2 fold change (MMSE): condition sort vs whole
## lfcSE               results           posterior SD: condition sort vs whole
## pvalue              results       Wald test p-value: condition sort vs whole
## padj                results                         BH adjusted p-values

summary(result)

##
## out of 1454 with nonzero total read count
## adjusted p-value < 0.05
## LFC > 0 (up)       : 31, 2.1%
## LFC < 0 (down)     : 35, 2.4%
## outliers [1]       : 0, 0%
## low counts [2]     : 986, 68%
## (mean count < 2)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results

### Results extraction

res_all_3 <- data.frame(result) %>% rownames_to_column(var = "circID")


res_all_3<- cbind(res_all_3, circCount_3[,2:5])

res_all_3<- mutate(res_all_3, ce11_loci = paste0("chr",Chr,":",Start-1,"-",End))

res_sig_3 <- subset(res_all_3, padj < 0.05)

#result export
write.table(res_all_3, "./circDE_sortwhole_3.txt", quote = F, row.names = F)

## Subset normalized counts to significant genes
sig_norm_counts <- normalized_circCounts[res_sig_3$circID, ]
## Choose a color palette from RColorBrewer
```

```
library(RColorBrewer)
heat_colors <- brewer.pal(6, "YlOrRd")

##Plot heatmap using pheatmap
pheatmap(sig_norm_counts,
         color = heat_colors,
         cluster_rows = T,
         show_rownames = F,
         annotation = select(metadata, condition),
         scale = "row")
```

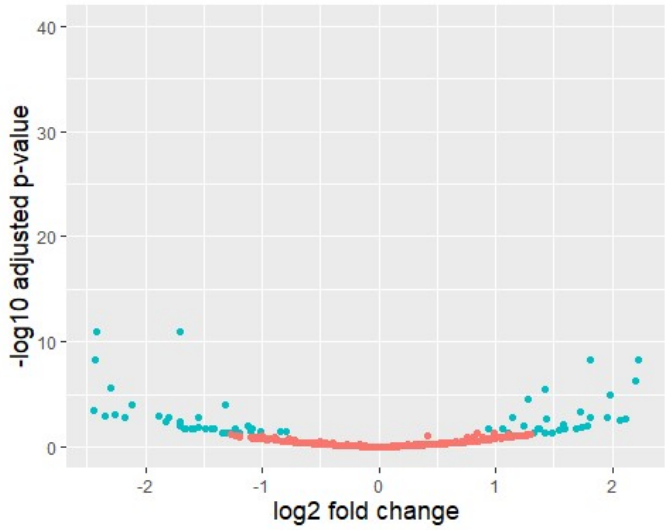



```
### Volcano plot

#Obtain logical vector regarding whether padj values are less than 0.05
res_all_3 <- mutate(res_all_3, threshold = padj<0.05) %>% mutate(updown = ifelse(threshold ==
F, 0, ifelse(log2FoldChange > 0, 1,2)))

ggplot(res_all_3) +
  geom_point(aes(x = log2FoldChange, y = -log10(padj), color = threshold)) +
  xlab("log2 fold change") +
  ylab("-log10 adjusted p-value") +
  theme(legend.position = "none",
        plot.title = element_text(size = rel(1.5), hjust = 0.5),
        axis.title = element_text(size = rel(1.25)))

## Warning: Removed 986 rows containing missing values (geom_point).
```



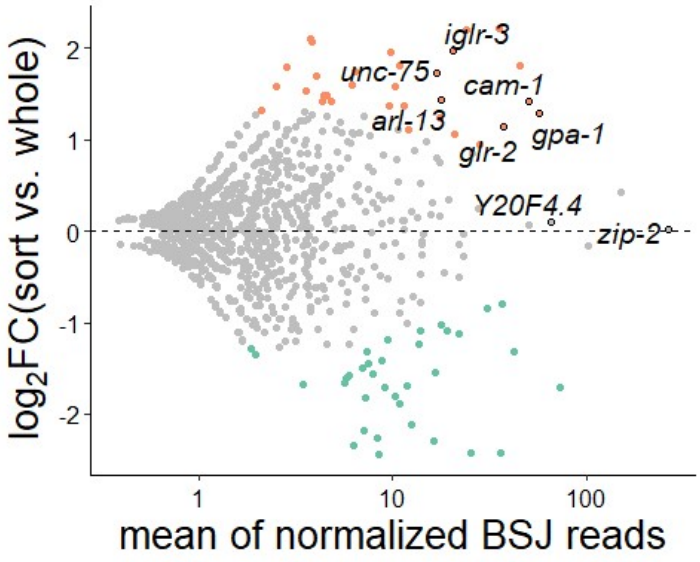

```
##adjust ylim
ggplot(res_all_3) +
  geom_point(aes(x = log2FoldChange, y = -log10(padj), color = threshold)) +
  xlab("log2 fold change") +
  ylab("-log10 adjusted p-value") +
  ylim(0, 40) +
  theme(legend.position = "none",
        plot.title = element_text(size = rel(1.5), hjust = 0.5),
        axis.title = element_text(size = rel(1.25)))
```

```
## Warning: Removed 986 rows containing missing values (geom_point).
```



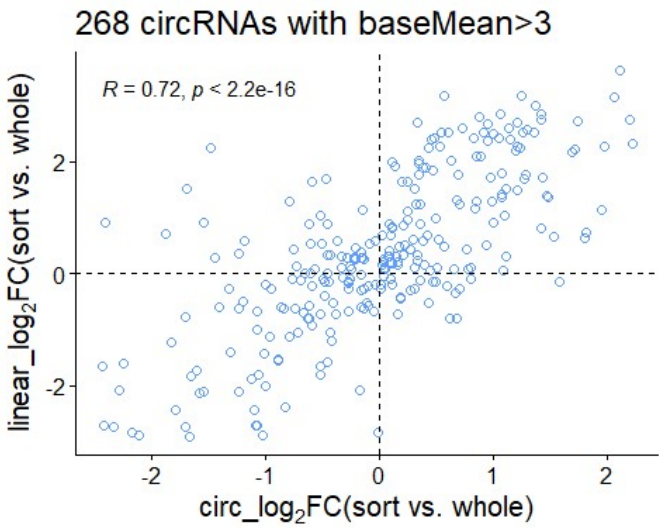

```
## highlight some genes
library(ggrepel)
#choose genes to label
gene <- c("glr-2", "gpa-1", "iglr-3","unc-75", "zip-2", "arl-13", "Y20F4.4", "unc-89", "sma-9", "nab-1",
"tbc-17","lin-1")

gene<-c("glr-2", "gpa-1", "iglr-3","unc-75", "zip-2", "arl-13", "Y20F4.4","cam-1")
subdata = filter(res_all_3, Gene  %in% gene) %>% filter(baseMean>10)
```

```
p3<- ggplot(res_all_3, aes(x = log2FoldChange, y = -log10(padj)), label = Gene) +
  geom_point(aes(color = threshold)) +
  geom_vline(xintercept = 0, lty = 8) +
  geom_hline(yintercept = 1, lty = 8) +
  geom_point(data = subdata, aes(x = log2FoldChange, y = -log10(padj)), color = "red") +
  geom_text_repel(data = subdata, size = 6, label = subdata$Gene, fontface = "italic") +
  scale_color_manual(values = c("grey", "#00BFC4")) +
  xlab("log2 fold change") +
  ylab("-log10 adjusted p-value") +
  theme_pubr()+
  theme(legend.position = "none",
        plot.title = element_text(size = rel(1.5), hjust = 0.5),
        axis.title = element_text(size = rel(1.7)))

## Draw MAplot by ggplot2

#replace_na is to convert NA values in threashold to FALSE

MA_circ_3 <- ggplot(replace_na(res_all_3, list(updown = FALSE)),
                    aes(x = baseMean, y = log2FoldChange), label = Gene) +
  geom_point(aes(color = as.factor(updown))) +
  scale_x_log10() +
  geom_hline(yintercept = 0, lty = 8) +
  geom_point(data = subdata, aes(x = baseMean, y = log2FoldChange), color = "black", shape =
1) +
  geom_text_repel(data = subdata, size = 5, label = subdata$Gene, fontface = "italic") +
  scale_color_manual(values = c("grey","#FC8D62", "#66C2A5")) +
  xlab("mean of normalized BSJ reads") +
  ylab(expression(log["2"]*"FC(sort vs. whole)")) +
  theme_pubr()+
  theme(legend.position = "none",
        plot.title = element_text(size = rel(1.5), hjust = 0.5),
        axis.title = element_text(size = rel(1.7)),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_rect(fill = "transparent",colour = NA),
        plot.background = element_rect(fill = "transparent",colour = NA)
        )

MA_circ_3
```

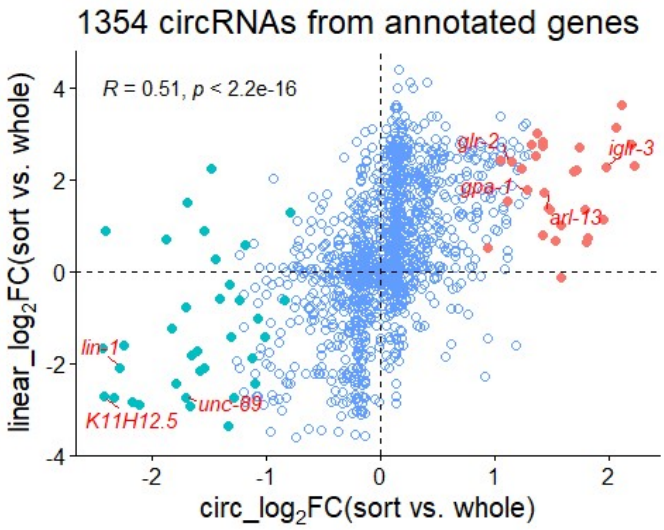


7.3.5 circDE vs linear DE

```
library(dplyr)

library(ggpubr)

#import circDE results from "DE analyisis of circRNAs using DESeq2.Rmd", filter out circRNAs
from not annotated loci

setwd("~/RNA-seq analysis/")
circ_DE <- read.table("./circDE/circDE_sortwhole_3.txt", header = T) %>% filter (Gene !=
"not_annotated") %>% dplyr::rename(gene_name = Gene)


#import linearDE results from "DE analysis using DESeq2.Rmd"

linear_DE <- read.csv("./linearDE.csv") %>% filter(gene_name %in% circ_DE$gene_name)

#merge 2 dataframes by gene_name

merged<- merge(circ_DE, linear_DE, by = "gene_name")

write.csv(merged, "circDE_linearDE.csv", quote = F, row.names = F)

##filter with baseMean.x>3

pCL<-ggscatter(filter(merged, baseMean.x>0), x = "log2FoldChange.x", y = "log2FoldChange.y",
size = 2, , fill = NA, shape = 21, color = "#619CFF", title = "1354 circRNAs from annotated genes")
+
  stat_cor(method = "pearson") +
  labs(x = expression(circ_log["2"]*"FC(sort vs. whole)"),
       y=expression(linear_log["2"]*"FC(sort vs. whole)" )) +
  geom_hline(yintercept = 0, linetype = "dashed") +
  geom_vline(xintercept = 0, linetype = "dashed") +
  theme(legend.position = "none",
        plot.title = element_text(size = rel(1.5), hjust = 0),
        axis.title = element_text(size = rel(1.25)))


pCL3<-ggscatter(filter(merged, baseMean.x>3), x = "log2FoldChange.x", y = "log2FoldChange.y",
size = 2, , fill = NA, shape = 21,  color = "#619CFF", title = "268 circRNAs with baseMean>3") +
  stat_cor(method = "pearson") +
  labs(x = expression(circ_log["2"]*"FC(sort vs. whole)"),
       y=expression(linear_log["2"]*"FC(sort vs. whole)" )) +
  geom_hline(yintercept = 0, linetype = "dashed") +
  geom_vline(xintercept = 0, linetype = "dashed") +
  theme(legend.position = "none",
        plot.title = element_text(size = rel(1.5), hjust = 0),
        axis.title = element_text(size = rel(1.25)))
pCL3
```

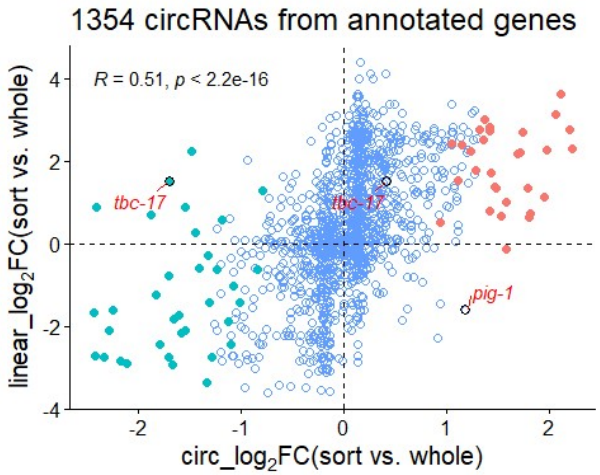



```
#label some genes

library(ggrepel)
gene <- c("glr-2", "gpa-1", "iglr-3", "arl-13", "gtl-2  ", "K11H12.5", "lin-1",  "unc-89" )
gene1<- c("tbc-17", "pig-1")
subdata = filter(merged, gene_name  %in% gene) %>% filter(baseMean.x >10)
subdata1= filter(merged, gene_name  %in% gene1)
#highlight significantly DE circRNAs
sig_circ<- filter(merged, padj.x <0.05) %>% mutate(updown.x = ifelse(log2FoldChange.x>0, 1, 2))

# No labeling
pCL + geom_point(data = sig_circ, aes(x = log2FoldChange.x , y = log2FoldChange.y, color =
as.factor(updown.x)), size = 2) + scale_fill_manual(values = c("#FC8D62", "#66C2A5")) +
  geom_text_repel(data = subdata, size = 4,color = "red", label = subdata$gene_name, fontface =
"italic",
                  point.padding = unit(0.25, "lines"),
                box.padding = unit(0.25, "lines"),
                min.segment.length = 0)
```

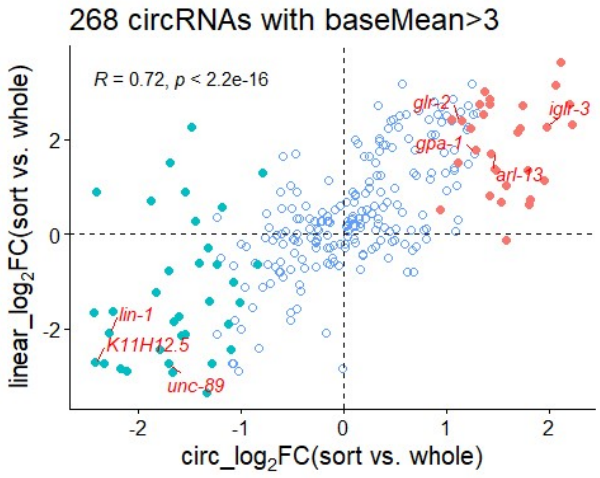



```
#label tbc-17 and pig-1
pCL +
```

```
 geom_point(data = sig_circ,
            aes(x = log2FoldChange.x , y = log2FoldChange.y, color = as.factor(updown.x)),
            size = 2) +
 scale_fill_manual(values = c("#FC8D62", "#66C2A5")) +
 geom_point(data = subdata1,
            aes(x = log2FoldChange.x , y = log2FoldChange.y),
            shape = 1, size = 2 , color = "black", stroke = 1) +
 geom_text_repel(data = subdata1, size = 4,color = "red", label = subdata1$gene_name, fontface
= "italic",
                  point.padding = unit(0.25, "lines"),
                 box.padding = unit(0.25, "lines"),
                 min.segment.length = 0)
```

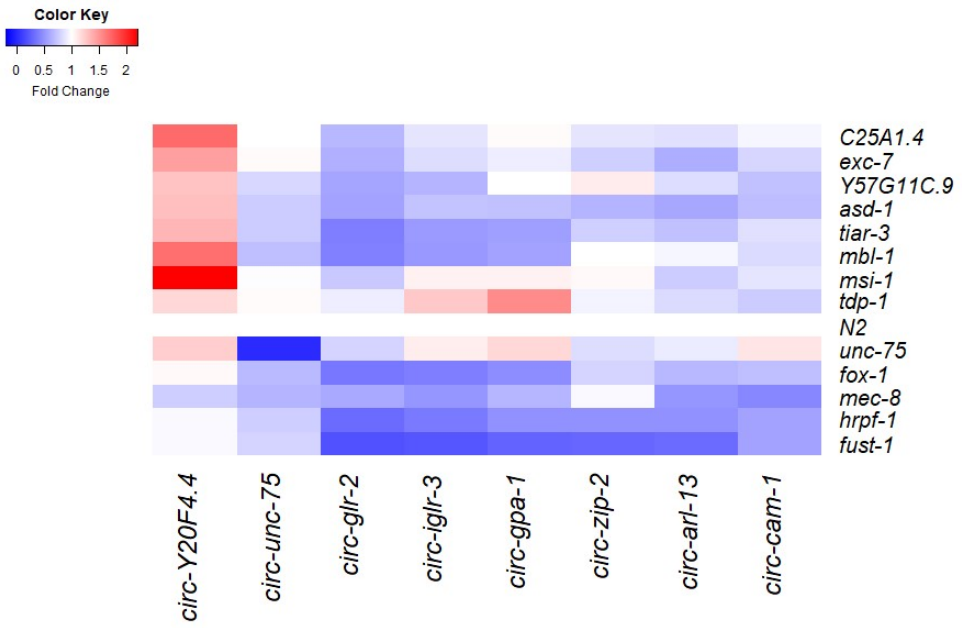



```
pCL3 + geom_point(data = sig_circ, aes(x = log2FoldChange.x , y = log2FoldChange.y, color =
as.factor(updown.x)), size = 2) + scale_fill_manual(values = c("#FC8D62", "#66C2A5")) +
 geom_text_repel(data = subdata, size = 4,color = "red", label = subdata$gene_name, fontface =
"italic",
                  point.padding = unit(0.25, "lines"),
                 box.padding = unit(0.25, "lines"),
                 min.segment.length = 0)
```



7.3.6 Heatplot in Figure 3.13A

```
## heatmap plot for circRNA fold change in RBP mutants

##import result data

library(readxl)
library(dplyr)

library(gplots)

##select(-5) to remove circ-nab-1
qPCR_RBP <- read_excel("./assaydata/qPCR_RBP.xlsx") %>% select(-5) %>%
tibble::column_to_rownames( var = "Sample")
df<- as.matrix(qPCR_RBP)
df

##          circ-glr-2  circ-unc-75 circ-gpa-1 circ-zip-2 circ-Y20F4.4 circ-cam-1
## asd-1      0.5629325 0.7564200759  0.7056788  0.6490386    1.3077674  0.6873996
## C25A1.4    0.6614507 0.9931094050  1.0233464  0.8764090    1.6980375  0.9513217
## exc-7      0.6209584 1.0250390768  0.9159197  0.7785423    1.4593503  0.8080969
## fox-1      0.3614715 0.6710472107  0.4681549  0.7978907    1.0357271  0.6954612
## fust-1     0.1864776 0.7965794802  0.2632979  0.2881918    0.9629277  0.5699885
## hrpf-1     0.3032248 0.7677495480  0.4812411  0.4875642    0.9689784  0.5665971
## mbl-1      0.4060341 0.6965837479  0.5667545  1.0038055    1.6789204  0.8228701
## mec-8      0.5964758 0.6441518664  0.6586121  0.9665455    0.7695940  0.4371135
## msi-1      0.7444817 0.9873660803  1.0662857  1.0379319    2.2009530  0.8733109
## tdp-1      0.9199061 1.0200546980  1.5474802  0.9433300    1.1806319  0.7566001
## tiar-3     0.3900975 0.7530415058  0.5452510  0.7709072    1.3460126  0.8563417
## unc-75     0.7942670 0.0001382712  1.1851201  0.8360450    1.2393516  1.1219867
## Y57G11C.9  0.5795485 0.8175869584  0.9954058  1.0906332    1.2850699  0.7013251
## N2         1.0000000 1.0000000000  1.0000000  1.0000000    1.0000000  1.0000000
##          circ-iglr-3 circ-arl-13
## asd-1      0.7122107   0.5804671
## C25A1.4    0.8731043   0.8531640
## exc-7      0.8302661   0.6134149
## fox-1      0.3936984   0.6655200
## fust-1     0.2011604   0.3070552
## hrpf-1     0.3709424   0.4826933
## mbl-1      0.5184013   0.9553524
## mec-8      0.5085844   0.5053511
## msi-1      1.0622433   0.7505279
## tdp-1      1.2507206   0.8299862
## tiar-3     0.5244303   0.7049152
## unc-75     1.0800723   0.9048862
## Y57G11C.9  0.6471557   0.8312083
## N2         1.0000000   1.0000000

heatmap.2(df, scale = "none", col = bluered(240), trace = "none", density.info = "none",
        breaks = seq(-0.2,2.2,0.01), margins = c(10,10), dendrogram = "none",
        lwid = c(2,8), lhei = c(2,8), key.xlab = "Fold Change",
        labCol=as.expression(lapply(colnames(df), function(a) bquote(italic(.(a))))),
        labRow=as.expression(lapply(rownames(df), function(a) bquote(italic(.(a))))),
        cexRow = 1.5,
        cexCol = 1.8
        )
```


Color Key
0 0.5 1 1.5 2
Fold Change
C25A1.4
exc-7
Y57G11C.9
asd-1
tiar-3
mbl-1
msi-1
tdp-1
N2
unc-75
fox-1
mec-8
hrpf-1
fust-1
circ-Y20F4.4
circ-unc-75
circ-glr-2
circ-iglr-3
circ-gpa-1
circ-zip-2
circ-arl-13
circ-cam-1

7.4 Plasmid sequences

fust-1p::fust-1::mRFP (9059 bp)

fust-1 genomic sequences (from promoter to before stop codon)

mRFP

Ampicillin resistance

AAGCttgcatgcctgcaggtcgactctagaggatcccc
TAGATCGGGCCCTCTTCGTACATTGCGTGGTATGAT
ATCAGCAGGGTATCATCCATTACTCCACATTGAGCCATTTTGTATCGTACCACCTTGTTTGCAT
CTTCTGATGGGATTGGTAAATGACTATATCTTGGCTTCACTTTTGGCACTTGCCAATCAAATCG
ACTTCCCTGGACAGGATCTTCCAGACTCTATGAACGAACAGAGAATGGAATTGGGAATAGTTC
AGAAAGAAGAAAAATTTTGCCAAAAAGTGGTTGAAAGTCTGCAAAACGCCCTTGTGCTCACGA
CGAAAATCAAGGAAGTTTATGAGAGTCGACTAAAGCAGAATAATATCAAACGTTCGTCGATATC
TTGTGGATCTCCAACATGCTGCGTCAGTGATATGCCAAAAAGCCAACAGGATCAAGATATCTT
CTTCTGCGTTGCTTGTCCATTTGCAATTCATATGGTGTGTGCTGGTAGATATACGGAGGATCAA
CGCCAGGCAGCTAAACATCATCCGACTGCTTGTTTAAAATGTCAAGAAAAACATATTCTTTCTA
ATGAGGAAATGATTGAACATGCTAATGAGGCACTTGATGAGCTTGCCAGGCGCTTACAACACG
AGACAAGCCATCTTAGTAACTTGAAAGCTGCTCGCCAAGACATGACAACGGCTATTTCGACTC
GAACTGGGCCAACAAAACAGAAACTCGAGGAAGTTTTGGATTATATTGGCTGTAGTCAAAAAA
TTAACTATCAGGGATTAACCGGAAATCAAGTAAATTTGTGATTTTCGGTATTTGAAAATATTCTA
ATCTTTAGGTTCGCCAACTACTAAGGACAGAAAACATTGACAAAGTACTAAAAGTCTTCACGGA
CTGTAGTTGGATCCAGAATATGAGAAACTTCATGAATGATTTGGCGTCACTCATGTCAGCAAG
CAATAACGCAGTCTTCACAGATGATGATTTGAATGATTTCGCTAAAACTTTGAAGAACTTGAAG
AGAAACATCAGAATTCTTCATCCTTCGATGGGAGTCACTCCTAAACTTCACATCCTCTGTACTC
ACCTCGAGCCTTACATTAGAGCAAAAAGAACTTGGGGTCGTACTTCAGAACAAGGGATGGAAG
CGTTTCATGTGCTCTTCAGGACAACAGAGACGCGATTTTCTTCTGTGATGTCTCTCAAGCTAC
GCGCTGAGTTGATGATCGACTATTTCGCTAACATCAACTATATTTCCGATAAAAGAAACTAGTG
ACAGCAAACATTTCTAGGAAACCGTTTTTATAGTACAtctctctctctctctctcactcACATGttttttttCGATTA
AATCAATTAAACGAGCAGAAAAGCCATGAATGATCTCTAAATTCCGACTTACAAAGTGTTTTGA
GTTGATGATAAGCGTGAATCCTAATTTTCAATTGCCTAAACTAACAAACGGGTTGTGAATGGGT
TCAGATTGGGGTACAGCATGAAAATatttcaaaagaagtgaatacaattttagaaaaaccgaatacaaGCAATGTA
AACATGaaaatgaaacaaaaaaacgaaaaaaaaaTCAACCTTGTTCGAAGAAAATGATGAGAATTAGAAT
GAGCAAGGAAATGCGAGAGACTTGGAAATGTGAGAGACGCAGCATTTCACGCGAAACGCGCC
AGCAATCGTCGCGCTTGCGCATTTCGAGACGAGGATACTGTGTAAAATTGTAATCATTTGTTG
CAATTTTTGTTCTTGAGTTAACTTTATTTTTCATTTTCCTCGACCTAATTTCTCAAAATTGTTAAA
ATGTATACTATTTATTACCCTGCAGTGCTCAAATCTTGTGATGAGCACTATCAAAGTTTTAGATG
GAGttttttaaaaagtatttttttttagtataattgttttttttcagtaaatttagttttttGGAGCAGAACGCTGAAATCTGCCAAGGT
TCGCTACAACCAGATATATCTAGATATAACTGCAAATGTGCGCGCTCCATTGAGACTACACGC
GACGCACAACCGCGACGCGTATTGGACAGCGTTCTCCGTCTTCTTTGAACAGCATCTGTCCTC
AACCATTCTCTTCTCGTTTCTCTTTTAATTTTAAGATATATTTGCTCTTCAAAGGATAAATTTCAG
TCAACGCATCGACACGTAGCGAGTGCCTCGACTTGAATATCGACAAAAATGGGTTAGTTTCTT
TTTTAATAGTCGTTTTCTTATTGGAATATCGAAATCCGAGTGTCCACTGAAAGTATTTTATTTAT
CCTCCTAACACCATTCATGTTTTATAAATTAGGTTAATGCGTGAATTTGTGGGAGCTCCCGAAT
ACAAATCACGGGCGTTGCCATTTCTGTTTCCCGGTTATTTGAAAACATTATCGATTTCAGCGGC
TTACGACCAAAGTCAACCCGACTACAGCACACCCGAAGGGCAGCAGGCGTATTGGGCCTATT
ATcaacagcaacaacaacaacaacCAGGAGGTCAACCTGATGTAAGTATATGAGATTTTGTGGGCTGA
TCGCCTCACTATCATGGCTTTGCTCAGAATTGTTCAGGAGCATAGTTCTGCTGTTGAATACCAC
CAGGCTCACAGACGTTTCATAGTTTAAATTTCAGCAGGATCCGTACGCCGCTGCTGCTTATGG
AGGCCACGACCAAGCACAACAACCGCAAAATCCATATGCACCACCACCACCAGGAGCGGATC
CATACGGCCAAGGATCCGGAGGCCAATCCGGAGGATCTGACCCCTACGGGCAAAGTAGAGG
TGGCGGCCGTGGAGGATTCGGAGGCAGTCGTGGAGGAGGTGGATATGATGGTGGACGTGGC
GGAAGTCGTggaggatacgacggaggacgcggaggttacggcggtgatcgcggaggaCGTGGTGGTGGTAGAGG
TTTGTTCTTGTTTGATTAATGTATTCATAATAAACCATGTTTTTCAGGTGGCTACGACGGAGAAC
GCCGAGGTGGGAGCCGATGGGATGACGGAAACTCTGATCGGTAGGTAAACATCTGAATGTTT
CGTTAGCAATATGCGGTATTTCCAAATTCGTTTGCTTAACCATCCAAAAATTGTTGTGTTGTGTT
GAAAATCGAAACTGCTACAGAGTTCAGAGTATTCAGTGTTCATATACAAAATCTATTCACTATTA
AAAACTTGCGAAGAGTAATGTATTGTGTCCGTCATTGTAAGTTTCATGAAAAAAGGTCAGTGTC

CGCAATGAAGCACAAATTCATTGTGCTTTATCGATTTGTAGGGCTATGGCCCGGTTTCCTTCAA
AATATTATGCAGTTCAGCTATTATTACGTGTGTTCGGTATCAAAACGAGTCAAGAGTGAAGAAC
AAATGAAGTCTACCAAACGACTTCGAGCGTGCAGGAGATGCTTCTGTCTGAAATCTCTCATGC
TTGAAAAATCTTAAAATCTAAATGATCGTGTTTGTGCGTACaaaaaaaaaTCAATTGAAGGTGTAT
TATTGCGCGGTCGAAATCTGAAAACATTTTACTGTGGTGTTAGCATGAGAGAAGAAGACAGAA
GACTCCATTCAAGAATGATGTTTTGCAGACAGGGCGGACCTCCAGGAGGAAGAGGAGGTTAT
CAAGGTATTGAATACGTAAATTGCTTTATCTTTCGTACCACAACGAGTTTATTACCAAAATTaaaa
aaaaaTGGGTATCCTACCGAGTTTATCATCGAAATTCAAAAACAAAAAACATTAAACGAATATAA
AAATACTTGTTCCTGTGCTTTATATTATTGACAGTTGGACGAAAAAATTGTTTGTAAAAGTTTTC
ATATTTTCATTGCAGACAGAGGACCCCGACGTGACGGACCACCAAGTGGAGGAGGATACGGA
GGCGGTGGTGCTGCTTCAGGAAACCGCGAATTCGGATCCGATGGGCGAGTTGAGTTGAAGG
AGACCGTCTTTGTGCAGGGAATTTCCACCACGTATGTTTTGTAATCGCCCAATTAAAAAATGTA
TCTATATTTCAGTGCAAATGAAGCCTACATCGCCGACGTCTTCAGTACCTGCGGAGATATTGC
AAAGAACGATCGCGGACCGAGAATCAAGATTTACACCGATCGCAACACCGGAGAACCAAAAG
GAGAATGCATGATAACATTCGTCGATGCTTCTGCAGCTCAACAAGCTATAACTATGTACAATGG
GCAACCATTCCCTGGCGGCTCAAGCCCGATGAGTATTTCACTAGCCAAGTTCCGCGCTGATG
CAGGAGGTGAACGAGGTGGTCGTGGTGGACGAGGCGGTTTCGGAGGTGGTCGTGGTGGTCC
TATGGGAGGACGTGGTGGCTTTGgcggtgatcgtggaggatatggcggcggcggtggtcgtggaggattcgatggtggtc
gcggtggtggcggtggattccgtggcggagaccgaggaggattccgaggtggtgatagaggaggcttccgcggaggagatcgtggag
gattccgaggcggcgaccgtggtggcgatcgtggaggattcagaggaggCCGTGGAGTAGGTGGCGGAAATGCTAA
TATGGAACAAAGGAAGAATGACTGGCCATGTGAGCAGTGCGGAAATAGCAACTTCGCTTTCAG
AAGAGTTAGTTAATTTTgttttttcaaaaaaaaaacacgggttttctTCTAGGAATGCAATCAATGCCAAGCCC
CGAGACCAGATGGAGGATCCGGAGGTGGTGGTGGCGAGCGACGAGGAGGACCACCAGGAG
GTGACCGATACCGTCCATATgggattggCCAAAGGACCCAAAGGTATGTTTCGAATGATACTAACA
TAACATAGAACATTTTCAGGAGGACCCTTGCTTGGAGGGTACCGGTAGAAAAAATGGCCTCCT
CCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCGTGAACGG
CCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGC
CAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAG
TTCCAGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCT
GTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTG
ACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCG
GCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTC
CACCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGATGAGGCTGAAG
CTGAAGGACGGCGGCCACTACGACGCCGAGGTCAAGACCACCTACATGGCCAAGAAGCCCG
TGCAGCTGCCCGGCGCCTACAAGACCGACATCAAGCTGGACATCACCTCCCACAACGAGGAC
TACACCATCGTGGAACAGTACGAGCGCGCCGAGGGCCGCCACTCCACCGGCGCCTAAGAAT
TCCAACTGAGCGCCGGTCGCTACCATTACCAACTTGTCTGGTGTCAAAAATAATAGGGGCCGC
TGTCATCAGAGTAAGTTTAAACTGAGTTCTACTAACTAACGAGTAATATTTAAATTTTCAGCTCT
CGCGCCCGTGCCTCTGACTTCTAAGTCCAATTACTCTTCAACATCCCTACATGCTCTTTCTCCC
TGTGCTCCCACCCCCTATTTTTGTTATTATCAAAAAAACTTCTTCTTAATTTCTTTGTTTTTTAGC
TTCTTTTAAGTCACCTCTAACAATGAAATTGTGTAGATTCAAAAATAGAATTAATTCGTAATAAA
AAGTCGAAAAAAATTGTGCTCCCTCCCCCCATTAATAATAATTCTATCCCAAAATCTACACAAT
GTTCTGTGTACACTTCTTATGTTTTTTTTACTTCTGATAAATTTTTTTTTGAAACATCATAGAAAAA
ACCGCACACAAAATACCTTATCATATGTTACGTTTCAGTTTATGACCGCAATTTTTATTTCTTCG
CACGTCTGGGCCTCTCATGACGTCAAATCATGCTCATCGTGAAAAAGTTTTGGAGTATTTTTGG
AATTTTTCAATCAAGTGAAAGTTTATGAAATTAATTTTCCTGCTTTTGCTTTTTGGGGGTTTCCC
CTATTGTTTGTCAAGAGTTTCGAGGACGGCGTTTTTCTTGCTAAAATCACAAGTATTGATGAGC
ACGATGCAAGAAAGATCGGAAGAAGGTTTGGGTTTGAGGCTCAGTGGAAGGTGAGTAGAAGT
TGATAATTTGAAAGTGGAGTAGTGTCTATGGGGTTTTTGCCTTAAATGACAGAATACATTCCCA
ATATACCAAACATAACTGTTTCCTACTAGTCGGCCGTACGGGCCCTTTCGTCTCGCGCGTTTC
GGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTA
AGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCG
GGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTG
AAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGGCCTTAAGGGCCTCGTGAT
ACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTC
GGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCT
CATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAA
CATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAG
AAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAA
CTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATG

AGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAA
CTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAG
CATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAAC
ACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCAC
AACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACC
AAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAA
CTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAG
TTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAG
CCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCG
TATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGC
TGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTT
TAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTC
ATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATC
AAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCAC
CGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTG
GCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACT
TCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTG
CCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCG
CAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACA
CCGAACTGAGATACCTACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAG
GCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCA
GGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCG
ATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTT
TACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTC
TGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCG
AGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCC
CGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCA
GTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTA
TGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCT
ATGACCATGATTACGCCAAGCTgtaagtttaaacatgatcttactaactaactattctcatttaaattttcagAGCTTAAAA
ATGGCTGAAATCACTCACAACGATGGATACGCTAACAACTTGGAAATGAAAT

fust-1 exon 5 splicing reporter:

rgef-1p::fust-1::GFP::mCherry::unc-54 3'UTR (10128 bp)

rgef-1 promoter

fust-1 exon 4 to exon 6

GFP

mCherry

Ampicillin resistance

cgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccg
gtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgcca
cctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggc
cttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgc
agccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggc
cgattcattaatgcagctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtgagttagctcactcatta
ggcaccccaggctttacactttatgcttccggctcgtatgttgtgtggaattgtgagcggataacaatttcacacaggaaacagctatgaccat
gattacgaattcgaatggccatgggacgtcgacgcttgcatgcctgcagcgtttccgataccccttatatcagcacacattcagtcatcata
ccacaaaatgttgtgatagtgataggcaaaactaccgtatttcctctattaatcttgcatgcaagactaattttcgattaacccgtaggggtgca
agactaatagagactgcaagactattagaggctgaaatactaattttcgtatgctcaataattttggaaattggcctatttttttgtagaaacttgat
accgtttaaacaaggaaaaatacacacttttaatatttcatcaataatttgaacgattttgtgattttaagttcaatttgcccaaaaaaagacaa
tttttctgacgttaccgcaaagtaatgctggtcaggcaaaaagagcggtgcaaaaatataagagactgcaatactaatagaggaaatacg
gtaattgaattttagtgaaacttgcgacagttttcctcatttttttgttattccggcatacgagtgtggcatacgagtgaggtcatctttgtttcttccgt
ttcttcatcgcttttgaaaaaaaatgttgaagaaaacttcgtagccgatagccggataatgtttggcacggcgttccaacatatatcattggaa
atttttaagatttttgcgaaaaatcacattcttcacgatgagaacacgttattgaaggaatataaatcaagaataacatatagttatatttctctat
tactttaacgttaaatatgagcaaatttgagcattttgattgcgatgaaaagcagaatcgagtcaactgaaatccgttcaaaaacctaagctc
ggttcgcccgaaagtcgatttttccaatagccgaacagcactgcttctttatttttccaggtatttttggggtgtaagaattatatcttctttgtttgggt
atttacgtttccgaattccctttcttgatctgtcgctatttttcagaaaccagaacagtttatccttttcatgagaacctgaatagctcaaaaaccc
gtttaatttcttttcatcatctcatttgaaactttcctagtcttcttaatgatttccaggctcctcctacatttgtatctcaagatgcatacacttttatttccg
ctccaattcgaatcatctacttctctttttttcttcaactttctacttttttccattctttttctttcaaatgttcgacgtcttcaagtgactgctcttcattctcctc
atttttctcgaatttcattcttgtctctttttttgctcaaaaatggaaaatgaaagtgacgtgagatttggacggcgggacacgggggcagtagaa
gcagcaaaaaggagagaaagaggacacaaataagaagaacgaattcaaaaataagcggagaggagctatttccgtcaattctacctc
cccaatcttcatcaattcggctcaattgaatgacgtcacagggagataaaccgtttggatgagcgccgtccatcactgacgccatcccgtttg
ggacaagaaaagagaaaaaagagcacaaagtttttggtgacggatcttgtcaatcatatgaaagttgttctgattgattgtcagtttttttccta
ctttttggattctatccacttctgaacttttgacaagtttcaaactttctgaatcattttctatgcattttcctggaattctttttatgtaaaatatgaaata
gaatgtttttgaattcaagttctgctttttttcttctttttgttctgttcgggcttgggtatgctttttttaaaaattattttgcacatcgaccaataagtgcgc
aacttataaaattaatttatttttgttaatttttgaaatacttgtattgctttaagtgatctgacctcgccctgagctttccacgtagttatcaaatacaa
atcctccaagggtaacgtacctatattactgatctttataataactttatcacctgtccagttccagaggaattctgttaagcttataggcacaga
aggagtcatttctgctggtttttaatgatccaaatctttatttcaagtaaaaaactgaacacttgcgaataaaactatcagattaaccattcacca
aaaatgtgtttgaatctaaaactttctcagtattccaaatatagaaataaataaccacgacattgctaaaatctgtctgaattgtgtactccttac
cgtgaagtaataatggatataatgaatcgtttgaaatgaatgatcagcacatttttggtgaaagatcacaaataaggaataagcgacggaa
aataaaacgattatttcggatcaaaaatttgttgaagatcatatacactcgagaccaagattattctagacaattttcaaattggcttctttgtttgc
aaatcttcataataatctgtgaagtttggaaatttgaatttttaatctttttttcatagattatagtttttatttctttgcaaaactatattaaaaaccgatgc
attgttttagggaaattaatgagccttttgttcaacactaaaaacaataaaattaaaatttttggcttcatcatttgaccttttttaagttcgaaaacttt
tctcgtatttctgaaccgccaatttttttcacacatctctagacttttggtgcccgttccagaaagttaagtaattgctattctaagaaagttctcaac
attgttttttagttctgattgaattctgatgttccaggaatatattttaaatttaatattctttgcactatttctatactaactaaataataaatagtctaagt
atgttcaatgagaacaaaaaagctctatctatatttttctatcctatttcatttatatccttttcattttgaactcaccagcttctcttcttcttcttgtccat
atacttcttaaaagtctcgaacttctctctctcactctccatcaatttctttatggctaccgcttgcgctctgccgcttccgaagaaagaaggcgtt
ggcagagctctcaattctatttttttctcgccgtaggctatcgatttctcagcttctcttctctcttcctcactccaccaccaccacttcgggcttcttctttt
tcttcattttcgtcttcttcttcatcctcttctttttttttcagttccccctcactcctcccctatatgcgcgtgcgagaggggtgcaaaagcagcgcgc
atccgagaattgaaacgagaaaacggagacgcagcagcagttcgtcctcagaaaaatagccagtaaaaagagaaaaagatagaga
gagacctatcgcattttatttttcaaattgcaattcctgcaattcatgtgtgcgtgtgcccattttcaattcttcccgttatttttcagaccattaccaacg
ttctttctagatcttgagacaatcttccttctgctcaatcgttcgtcgtgaagacggcatcgacgacgacgacgatcagcagtaaaggatcatc
ggggacaagtttgtacaaaaaagcaggctaccATGGATTACAAGGATGACGATGACAAGGGGGTACCTGCC
CCAAAAAAAAAACGCAAAGTGGAGGACCCAGTACCCGGATCTGAATTCATATATctcgagGTTAA
CTACGCTACGACGGAGAACGCCGAGGTGGGAGCCGATGGGATGACGGAAACTCTGATCGGT
AGGTAAACATCTGAATGTTTCGTTAGCAATATGCGGTATTTCCAAATTCGTTTGCTTAACCATC

CAAAAATTGTTGTGTTGTGTTGAAAATCGAAACTGCTACAGAGTTCAGAGTATTCAGTGTTCAT
ATACAAAATCTATTCACTATTAAAAACTTGCGAAGAGTAATGTATTGTGTCCGTCATTGTAAGTT
TCATGAAAAAAGGTCAGTGTCCGCAATGAAGCACAAATTCATTGTGCTTTATCGATTTGTAGGG
CTATGGCCCGGTTTCCTTCAAAATATTATGCAGTTCAGCTATTATTACGTGTGTTCGGTATCAA
AACGAGTCAAGAGTGAAGAACAAATGAAGTCTACCAAACGACTTCGAGCGTGCAGGAGATGC
TTCTGTCTGAAATCTCTCATGCTTGAAAAATCTTAAAATCTAAATGATCGTGTTTGTGCGTACaa
aaaaaaaTCAATTGAAGGTGTATTATTGCGCGGTCGAAATCTGAAAACATTTTACTGTGGTGTTA
GCATGAGAGAAGAAGACAGAAGACTCCATTCAAGAATGATGTTTTGCAGACAGGGCGGACCT
CCAGGAGGAAGAGGAGGTTATCAAGGTATTGAATACGTAAATTGCTTTATCTTTCGTACCACAA
CGAGTTTATTACCAAAATTaaaaaaaaaaTGGGTATCCTACCGAGTTTATCATCGAAATTCAAAAAC
AAAAAACATTAAACGAATATAAAAATACTTGTTCCTGTGCTTTATATTATTGACAGTTGGACGAA
AAAATTGTTTGTAAAAGTTTTCATATTTTCATTGCAGACAGAGGACCCCGACGTGACGGACCAC
CAAGTGGAGGAGGATACGGAGGCGGTGGTGCTGCTTCAGGAAACCGCGAATTCGGATCCGA
TGGGCGAGTTGAGTTGGTATgcggccgcagtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgag
ctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttc
atctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgac
cacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaacta
caagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggca
acatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtga
acttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccccatcggcgacggcccc
gtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctgg
agttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagtaaaccgcggtgtgagcaagggcgaggaggataacatg
gccatcatcaaggagttcatgcgcttcaaggtgcacatggagggctccgtgaacggccacgagttcgagatcgagggcgagggcgagg
gccgcccctacgagggcacccagaccgccaagctgaaggtgaccaagggtggcccccctgcccttcgcctgggacatcctgtcccctca
gttcatgtacggctccaaggcctacgtgaagcaccccgccgacatccccgactacttgaagctgtccttccccgagggcttcaagtgggag
cgcgtgatgaacttcgaggacggcggcgtggtgaccgtgacccaggactcctccctgcaggacggcgagttcatctacaaggtgaagct
gcgcggcaccaacttcccctccgacggccccgtaatgcagaagaagaccatgggctgggaggcctcctccgagcggatgtaccccga
ggacggcgccctgaagggcgagatcaagcagaggctgaagctgaaggacggcggccactacgacgctgaggtcaagaccacctac
aaggccaagaagcccgtgcagctgcccggcgcctacaacgtcaacatcaagttggacatcacctcccacaacgaggactacaccatc
gtggaacagtacgaacgcgccgagggccgccactccaccggcggcatggacgagctgtacaagtaaacccagctttcttgtacaaagt
ggtccccgatgatcccccgggctgcaggaattcgatatgcagaagctgatctcagaggaggacctgcttcagaagcttatctcagaggag
gaccttcgacctcgaggggggcccggtaccatggtattgatatctgagctccgcatcggccgctgtcatcagatcgccatctcgcgcccgt
gcctctgacttctaagtccaattactcttcaacatccctacatgctctttctccctgtgctcccacccccctatttttgttattatcaaaaaaacttcttct
taatttctttgtttttttagcttcttttaagtcacctctaacaatgaaattgtgtagattcaaaaatagaattaattcgtaataaaaagtcgaaaaaaat
tgtgctccctccccccattaataataattctatcccaaaatctacacaatgttctgtgtacacttcttatgtttttttttacttctgataaatttttttttgaaa
catcatagaaaaaaccgcacacaaaataccttatcatatgttacgtttcagtttatgaccgcaatttttatttcttcgcacgtctgggcctctcatg
acgtcaaatcatgctcatcgtgaaaaagttttggagtatttttggaatttttcaatcaagtgaaagtttatgaaattaattttcctgcttttgctttttgg
gggtttcccctattgtttgtcaagagtttcgaggacggcgtttttcttgctaaaatcacaagtattgatgagcacgatgcaagaaagatcggaa
gaaggtttgggtttgaggctcagtggaaggtgagtagaagttgataatttgaaagtggagtagtgtctatggggtttttgccttaaatgacaga
atacattcccaatataccaaacataactgtttcctactagtggccacgtgggccgtgcaccttaagcttggcactggccgtcgttttacaacgt
cgtgactgggaaaaccctggcgttacccaacttaatcgccttgcagcacatccccctttcgccagctggcgtaatagcgaagaggcccgc
accgatcgcccttcccaacagttgcgcagcctgaatggcgaatggcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacacc
gcatacgtcaaagcaaccatagtacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacac
ttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccc
tttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgatttgggtgatggttcacgtagtgggccatcgccctgatagacg
gtttttcgcccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcgggctattcttttgatt
tataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaa
ttttatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagccccgacacccgccaacacccgctgacgcgccctgacg
ggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaac
gcgcgagacgaaagggcctcgtgatacgcctatttttataggttaatgtcatgataataatggtttcttagacgtcaggtggcacttttcgggga
aatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatatt
gaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattcccttttttgcggcattttgccttcctgtttttgctcacccagaaacgctg
gtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttc
gccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcg
gtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaatta
tgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgc
acaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgat
gcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggag
gcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgc
ggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacga

aatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaa

fust-1p::fust-1a(cDNA)::mRFP (7882 bp)

isoform a cDNA

mRFP

Ampicillin resistance

AAGCttgcatgcctgcaggtcgactctagaggatccccTAGATCGGGCCCTCTTCGTACATTGCGTGGTATGAT
ATCAGCAGGGTATCATCCATTACTCCACATTGAGCCATTTTGTATCGTACCACCTTGTTTGCAT
CTTCTGATGGGATTGGTAAATGACTATATCTTGGCTTCACTTTTGGCACTTGCCAATCAAATCG
ACTTCCCTGGACAGGATCTTCCAGACTCTATGAACGAACAGAGAATGGAATTGGGAATAGTTC
AGAAAGAAGAAAAATTTTGCCAAAAAGTGGTTGAAAGTCTGCAAAACGCCCTTGTGCTCACGA
CGAAAATCAAGGAAGTTTATGAGAGTCGACTAAAGCAGAATAATATCAAACGTTCGTCGATATC
TTGTGGATCTCCAACATGCTGCGTCAGTGATATGCCAAAAAGCCAACAGGATCAAGATATCTT
CTTCTGCGTTGCTTGTCCATTTGCAATTCATATGGTGTGTGCTGGTAGATATACGGAGGATCAA
CGCCAGGCAGCTAAACATCATCCGACTGCTTGTTTAAAATGTCAAGAAAAACATATTCTTTCTA
ATGAGGAAATGATTGAACATGCTAATGAGGCACTTGATGAGCTTGCCAGGCGCTTACAACACG
AGACAAGCCATCTTAGTAACTTGAAAGCTGCTCGCCAAGACATGACAACGGCTATTTCGACTC
GAACTGGGCCAACAAAACAGAAACTCGAGGAAGTTTTGGATTATATTGGCTGTAGTCAAAAAA
TTAACTATCAGGGATTAACCGGAAATCAAGTAAATTTGTGATTTTCGGTATTTGAAAATATTCTA
ATCTTTAGGTTCGCCAACTACTAAGGACAGAAAACATTGACAAAGTACTAAAAGTCTTCACGGA
CTGTAGTTGGATCCAGAATATGAGAAACTTCATGAATGATTTGGCGTCACTCATGTCAGCAAG
CAATAACGCAGTCTTCACAGATGATGATTTGAATGATTTCGCTAAAACTTTGAAGAACTTGAAG
AGAAACATCAGAATTCTTCATCCTTCGATGGGAGTCACTCCTAAACTTCACATCCTCTGTACTC
ACCTCGAGCCTTACATTAGAGCAAAAAGAACTTGGGGTCGTACTTCAGAACAAGGGATGGAAG
CGTTTCATGTGCTCTTCAGGACAACAGAGACGCGATTTTCTTCTGTGATGTCTCTCAAGCTAC
GCGCTGAGTTGATGATCGACTATTTCGCTAACATCAACTATATTTCCGATAAAAGAAACTAGTG
ACAGCAAACATTTCTAGGAAACCGTTTTTATAGTACAtctctctctctctctctcactcACATGttttttttCGATTA
AATCAATTAAACGAGCAGAAAAGCCATGAATGATCTCTAAATTCCGACTTACAAAGTGTTTTGA
GTTGATGATAAGCGTGAATCCTAATTTTCAATTGCCTAAACTAACAAACGGGTTGTGAATGGGT
TCAGATTGGGGTACAGCATGAAAATatttcaaaagaagtgaatacaattttagaaaaaccgaatacaaGCAATGTA
AACATGaaaatgaaacaaaaaaacgaaaaaaaaaTCAACCTTGTTCGAAGAAAATGATGAGAATTAGAAT
GAGCAAGGAAATGCGAGAGACTTGGAAATGTGAGAGACGCAGCATTTCACGCGAAACGCGCC
AGCAATCGTCGCGCTTGCGCATTTCGAGACGAGGATACTGTGTAAAATTGTAATCATTTGTTG
CAATTTTTGTTCTTGAGTTAACTTTATTTTTCATTTTCCTCGACCTAATTTCTCAAAATTGTTAAA
ATGTATACTATTTATTACCCTGCAGTGCTCAAATCTTGTGATGAGCACTATCAAAGTTTTAGATG
GAGttttttaaaaagtatttttttttagtataattgttttttttcagtaaatttagttttttGGAGCAGAACGCTGAAATCTGCCAAGGT
TCGCTACAACCAGATATATCTAGATATAACTGCAAATGTGCGCGCTCCATTGAGACTACACGC
GACGCACAACCGCGACGCGTATTGGACAGCGTTCTCCGTCTTCTTTGAACAGCATCTGTCCTC
AACCATTCTCTTCTCGTTTCTCTTTTAATTTTAAGATATATTTGCTCTTCAAAGGATAAATTTCAG
TCAACGCATCGACACGTAGCGAGTGCCTCGACTTGAATATCGACAAAAATGGCGGCTTACGA
CCAAAGTCAACCCGACTACAGCACACCCGAAGGGCAGCAGGCGTATTGGGCCTATTATcaaca
gcaacaacaacaacaacCAGGAGGTCAACCTGATCAGGATCCGTACGCCGCTGCTGCTTATGGAGG
CCACGACCAAGCACAACAACCGCAAAATCCATATGCACCACCACCACCAGGAGCGGATCCAT
ACGGCCAAGGATCCGGAGGCCAATCCGGAGGATCTGACCCCTACGGGCAAAGTAGAGGTGG
CGGCCGTGGAGGATTCGGAGGCAGTCGTGGAGGAGGTGGATATGATGGTGGACGTGGCGG
AAGTCGTggaggatacgacggaggacgcggaggttacggcggtgatcgcggaggaCGTGGTGGTGGTAGAGGTG
GCTACGACGGAGAACGCCGAGGTGGGAGCCGATGGGATGACGGAAACTCTGATCGACAGGG
CGGACCTCCAGGAGGAAGAGGAGGTTATCAAGACAGAGGACCCCGACGTGACGGACCACCA
AGTGGAGGAGGATACGGAGGCGGTGGTGCTGCTTCAGGAAACCGCGAATTCGGATCCGATG
GGCGAGTTGAGTTGAAGGAGACCGTCTTTGTGCAGGGAATTTCCACCACTGCAAATGAAGCC
TACATCGCCGACGTCTTCAGTACCTGCGGAGATATTGCAAAGAACGATCGCGGACCGAGAAT
CAAGATTTACACCGATCGCAACACCGGAGAACCAAAAGGAGAATGCATGATAACATTCGTCGA
TGCTTCTGCAGCTCAACAAGCTATAACTATGTACAATGGGCAACCATTCCCTGGCGGCTCAAG
CCCGATGAGTATTTCACTAGCCAAGTTCCGCGCTGATGCAGGAGGTGAACGAGGTGGTCGTG
GTGGACGAGGCGGTTTCGGAGGTGGTCGTGGTGGTCCTATGGGAGGACGTGGTGGCTTTGg
cggtgatcgtggaggatatggcggcggcggtggtcgtggaggattcgatggtggtcgcggtggtggcggtggattccgtggcggagaccg
aggaggattccgaggtggtgatagaggaggcttccgcggaggagatcgtggaggattccgaggcggcgaccgtggtggcgatcgtgga
ggattcagaggaggCCGTGGAGTAGGTGGCGGAAATGCTAATATGGAACAAAGGAAGAATGACTGG
CCATGTGAGCAGTGCGGAAATAGCAACTTCGCTTTCAGAAGAGAATGCAATCAATGCCAAGCC

CCGAGACCAGATGGAGGATCCGGAGGTGGTGGTGGCGAGCGACGAGGAGGACCACCAGGA
GGTGACCGATACCGTCCATATgggattggCCAAAGGACCCAAAGGTATGTTTCGAATGATACTAA
CATAACATAGAACATTTTCAGGAGGACCCTTGCTTGGAGGGTACCGGTAGAAAAAATGGCCTC
CTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCGTGAACG
GCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCG
CCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCA
GTTCCAGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGC
TGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGT
GACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGC
GGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCT
CCACCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGATGAGGCTGAA
GCTGAAGGACGGCGGCCACTACGACGCCGAGGTCAAGACCACCTACATGGCCAAGAAGCCC
GTGCAGCTGCCCGGCGCCTACAAGACCGACATCAAGCTGGACATCACCTCCCACAACGAGGA
CTACACCATCGTGGAACAGTACGAGCGCGCCGAGGGCCGCCACTCCACCGGCGCCTAAGAA
TTCCAACTGAGCGCCGGTCGCTACCATTACCAACTTGTCTGGTGTCAAAAATAATAGGGGCCG
CTGTCATCAGAGTAAGTTTAAACTGAGTTCTACTAACTAACGAGTAATATTTAAATTTTCAGCTC
TCGCGCCCGTGCCTCTGACTTCTAAGTCCAATTACTCTTCAACATCCCTACATGCTCTTTCTCC
CTGTGCTCCCACCCCCTATTTTTGTTATTATCAAAAAAACTTCTTCTTAATTTCTTTGTTTTTTAG
CTTCTTTTAAGTCACCTCTAACAATGAAATTGTGTAGATTCAAAAATAGAATTAATTCGTAATAA
AAAGTCGAAAAAAATTGTGCTCCCTCCCCCCATTAATAATAATTCTATCCCAAAATCTACACAAT
GTTCTGTGTACACTTCTTATGTTTTTTTTACTTCTGATAAATTTTTTTTGAAACATCATAGAAAAA
ACCGCACACAAAATACCTTATCATATGTTACGTTTCAGTTTATGACCGCAATTTTTATTTCTTCG
CACGTCTGGGCCTCTCATGACGTCAAATCATGCTCATCGTGAAAAAGTTTTGGAGTATTTTTGG
AATTTTTCAATCAAGTGAAAGTTTATGAAATTAATTTTCCTGCTTTTGCTTTTTGGGGGTTTCCC
CTATTGTTTGTCAAGAGTTTCGAGGACGGCGTTTTTCTTGCTAAAATCACAAGTATTGATGAGC
ACGATGCAAGAAAGATCGGAAGAAGGTTTGGGTTTGAGGCTCAGTGGAAGGTGAGTAGAAGT
TGATAATTTGAAAGTGGAGTAGTGTCTATGGGGTTTTTGCCTTAAATGACAGAATACATTCCCA
ATATACCAAACATAACTGTTTCCTACTAGTCGGCCGTACGGGCCCTTTCGTCTCGCGCGTTTC
GGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTA
AGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCG
GGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTG
AAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGGCCTTAAGGGCCTCGTGAT
ACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTC
GGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCT
CATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAA
CATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAG
AAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAA
CTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATG
AGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAA
CTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAG
CATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAAC
ACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCAC
AACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACC
AAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAA
CTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAG
TTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAG
CCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCG
TATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGC
TGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTT
TAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTC
ATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATC
AAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCAC
CGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTG
GCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACT
TCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTG
CCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCG
CAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACA
CCGAACTGAGATACCTACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAG
GCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCA
GGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCG

ATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTT
TACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTC
TGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCG
AGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCC
CGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCA
GTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTA
TGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCT
ATGACCATGATTACGCCAAGCTgtaagtttaaacatgatcttactaactaactattctcatttaaattttcagAGCTTAAAA
ATGGCTGAAATCACTCACAACGATGGATACGCTAACAACTTGGAAATGAAAT

fust-1p::fust-1b(cDNA)::mRFP (7648 bp)

isoform b cDNA

mRFP

Ampicillin resistance

AAGCttgcatgcctgcaggtcgactctagaggatccccTAGATCGGGCCCTCTTCGTACATTGCGTGGTATGAT
ATCAGCAGGGTATCATCCATTACTCCACATTGAGCCATTTTGTATCGTACCACCTTGTTTGCAT
CTTCTGATGGGATTGGTAAATGACTATATCTTGGCTTCACTTTTGGCACTTGCCAATCAAATCG
ACTTCCCTGGACAGGATCTTCCAGACTCTATGAACGAACAGAGAATGGAATTGGGAATAGTTC
AGAAAGAAGAAAAATTTTGCCAAAAAGTGGTTGAAAGTCTGCAAAACGCCCTTGTGCTCACGA
CGAAAATCAAGGAAGTTTATGAGAGTCGACTAAAGCAGAATAATATCAAACGTTCGTCGATATC
TTGTGGATCTCCAACATGCTGCGTCAGTGATATGCCAAAAAGCCAACAGGATCAAGATATCTT
CTTCTGCGTTGCTTGTCCATTTGCAATTCATATGGTGTGTGCTGGTAGATATACGGAGGATCAA
CGCCAGGCAGCTAAACATCATCCGACTGCTTGTTTAAAATGTCAAGAAAAACATATTCTTTCTA
ATGAGGAAATGATTGAACATGCTAATGAGGCACTTGATGAGCTTGCCAGGCGCTTACAACACG
AGACAAGCCATCTTAGTAACTTGAAAGCTGCTCGCCAAGACATGACAACGGCTATTTCGACTC
GAACTGGGCCAACAAAACAGAAACTCGAGGAAGTTTTGGATTATATTGGCTGTAGTCAAAAAA
TTAACTATCAGGGATTAACCGGAAATCAAGTAAATTTGTGATTTTCGGTATTTGAAAATATTCTA
ATCTTTAGGTTCGCCAACTACTAAGGACAGAAAACATTGACAAAGTACTAAAAGTCTTCACGGA
CTGTAGTTGGATCCAGAATATGAGAAACTTCATGAATGATTTGGCGTCACTCATGTCAGCAAG
CAATAACGCAGTCTTCACAGATGATGATTTGAATGATTTCGCTAAAACTTTGAAGAACTTGAAG
AGAAACATCAGAATTCTTCATCCTTCGATGGGAGTCACTCCTAAACTTCACATCCTCTGTACTC
ACCTCGAGCCTTACATTAGAGCAAAAAGAACTTGGGGTCGTACTTCAGAACAAGGGATGGAAG
CGTTTCATGTGCTCTTCAGGACAACAGAGACGCGATTTTCTTCTGTGATGTCTCTCAAGCTAC
GCGCTGAGTTGATGATCGACTATTTCGCTAACATCAACTATATTTCCGATAAAAGAAACTAGTG
ACAGCAAACATTTCTAGGAAACCGTTTTTATAGTACAtctctctctctctctctcactcACATGttttttttCGATTA
AATCAATTAAACGAGCAGAAAAGCCATGAATGATCTCTAAATTCCGACTTACAAAGTGTTTTGA
GTTGATGATAAGCGTGAATCCTAATTTTCAATTGCCTAAACTAACAAACGGGTTGTGAATGGGT
TCAGATTGGGGTACAGCATGAAAATatttcaaaagaagtgaatacaattttagaaaaaccgaatacaaGCAATGTA
AACATGaaaatgaaacaaaaaaacgaaaaaaaaTCAACCTTGTTCGAAGAAAATGATGAGAATTAGAAT
GAGCAAGGAAATGCGAGAGACTTGGAAATGTGAGAGACGCAGCATTTCACGCGAAACGCGCC
AGCAATCGTCGCGCTTGCGCATTTCGAGACGAGGATACTGTGTAAAATTGTAATCATTTGTTG
CAATTTTTGTTCTTGAGTTAACTTTATTTTTCATTTTCCTCGACCTAATTTCTCAAAATTGTTAAA
ATGTATACTATTTATTACCCTGCAGTGCTCAAATCTTGTGATGAGCACTATCAAAGTTTTAGATG
GAGttttttaaaaagtatttttttttagtataattgttttttttcagtaaatttagtttttttGGAGCAGAACGCTGAAATCTGCCAAGGT
TCGCTACAACCAGATATATCTAGATATAACTGCAAATGTGCGCGCTCCATTGAGACTACACGC
GACGCACAACCGCGACGCGTATTGGACAGCGTTCTCCGTCTTCTTTGAACAGCATCTGTCCTC
AACCATTCTCTTCTCGTTTCTCTTTTAATTTTAAGATATATTTGCTCTTCAAAGGATAAATTTCAG
TCAACGCATCGACACGTAGCGAGTGCCTCGACTTGAATATCGACAAAAATGGAGGCCACGAC
CAAGCACAACAACCGCAAAATCCATATGCACCACCACCACCAGGAGCGGATCCATACGGCCA
AGGATCCGGAGGCCAATCCGGAGGATCTGACCCCTACGGGCAAAGTAGAGGTGGCGGCCGT
GGAGGATTCGGAGGCAGTCGTGGAGGAGGTGGATATGATGGTGGACGTGGCGGAAGTCGTg
gaggatacgacggaggacgcggaggttacggcggtgatcgcggaggaCGTGGTGGTGGTAGAGGTGGCTACGAC
GGAGAACGCCGAGGTGGGAGCCGATGGGATGACGGAAACTCTGATCGACAGAGGACCCCGA
CGTGACGGACCACCAAGTGGAGGAGGATACGGAGGCGGTGGTGCTGCTTCAGGAAACCGCG
AATTCGGATCCGATGGGCGAGTTGAGTTGAAGGAGACCGTCTTTGTGCAGGGAATTTCCACC
ACTGCAAATGAAGCCTACATCGCCGACGTCTTCAGTACCTGCGGAGATATTGCAAAGAACGAT
CGCGGACCGAGAATCAAGATTTACACCGATCGCAACACCGGAGAACCAAAAGGAGAATGCAT
GATAACATTCGTCGATGCTTCTGCAGCTCAACAAGCTATAACTATGTACAATGGGCAACCATTC
CCTGGCGGCTCAAGCCCGATGAGTATTTCACTAGCCAAGTTCCGCGCTGATGCAGGAGGTGA
ACGAGGTGGTCGTGGTGGACGAGGCGGTTTCGGAGGTGGTCGTGGTGGTCCTATGGGAGGA
CGTGGTGGCTTTGgcggtgatcgtggaggatatggcggcggcggtggtcgtggaggattcgatggtggtcgcggtggtggcggt
ggattccgtggcggagaccgaggaggattccgaggtggtgatagaggaggcttccgcggaggagatcgtggaggattccgaggcggc
gaccgtggtggcgatcgtggaggattcagaggaggCCGTGGAGTAGGTGGCGGAAATGCTAATATGGAACAA
AGGAAGAATGACTGGCCATGTGAGCAGTGCGGAAATAGCAACTTCGCTTTCAGAAGAGAATG
CAATCAATGCCAAGCCCCGAGACCAGATGGAGGATCCGGAGGTGGTGGTGGCGAGCGACGA
GGAGGACCACCAGGAGGTGACCGATACCGTCCATATgggattggCCAAAGGACCCAAAGGTATG
TTTCGAATGATACTAACATAACATAGAACATTTTCAGGAGGACCCTTGCTTGGAGGGTACCGG

TAGAAAAAATGGCCTCCTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATG
GAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTAC
GAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGG
GACATCCTGTCCCCTCAGTTCCAGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACAT
CCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCG
AGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTA
CAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC
ATGGGCTGGGAGGCCTCCACCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAG
ATCAAGATGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCCGAGGTCAAGACCACCT
ACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAAGACCGACATCAAGCTGGACAT
CACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCGCCGAGGGCCGCCAC
TCCACCGGCGCCTAAGAATTCCAACTGAGCGCCGGTCGCTACCATTACCAACTTGTCTGGTGT
CAAAAATAATAGGGGCCGCTGTCATCAGAGTAAGTTTAAACTGAGTTCTACTAACTAACGAGTA
ATATTTAAATTTTCAGCTCTCGCGCCCGTGCCTCTGACTTCTAAGTCCAATTACTCTTCAACAT
CCCTACATGCTCTTTCTCCCTGTGCTCCCACCCCCTATTTTTGTTATTATCAAAAAAACTTCTTC
TTAATTTCTTTGTTTTTTAGCTTCTTTTAAGTCACCTCTAACAATGAAATTGTGTAGATTCAAAAA
TAGAATTAATTCGTAATAAAAAGTCGAAAAAAATTGTGCTCCCTCCCCCCATTAATAATAATTCT
ATCCCAAAATCTACACAATGTTCTGTGTACACTTCTTATGTTTTTTTTACTTCTGATAAATTTTTT
TTGAAACATCATAGAAAAAACCGCACACAAAATACCTTATCATATGTTACGTTTCAGTTTATGAC
CGCAATTTTTATTTCTTCGCACGTCTGGGCCTCTCATGACGTCAAATCATGCTCATCGTGAAAA
AGTTTTGGAGTATTTTTGGAATTTTTCAATCAAGTGAAAGTTTATGAAATTAATTTTCCTGCTTTT
GCTTTTTGGGGGTTTCCCCTATTGTTTGTCAAGAGTTTCGAGGACGGCGTTTTTCTTGCTAAAA
TCACAAGTATTGATGAGCACGATGCAAGAAAGATCGGAAGAAGGTTTGGGTTTGAGGCTCAGT
GGAAGGTGAGTAGAAGTTGATAATTTGAAAGTGGAGTAGTGTCTATGGGGTTTTTGCCTTAAA
TGACAGAATACATTCCCAATATACCAAACATAACTGTTTCCTACTAGTCGGCCGTACGGGCCC
TTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGAC
GGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCG
GGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGT
GCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGGC
CTTAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGA
CGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACA
TTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGG
AAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTC
CTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCAC
GAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAA
GAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTG
ACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTAC
TCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCC
ATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGA
GCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGA
GCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAA
CGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTG
GATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTA
TTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCA
GATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGA
ACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCA
AGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAA
GATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCA
GACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCT
TGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTC
TTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGC
CGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCC
TGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGA
TAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCT
TGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAAGCGCCACG
CTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGC
GCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCAC
CTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGC
CAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCT
GCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGC

CGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATA
CGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCC
CGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCAC
CCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAAT
TTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTgtaagtttaaacatgatcttactaactaactattctc
atttaaattttcagAGCTTAAAAATGGCTGAAATCACTCACAACGATGGATACGCTAACAACTTGGAAA
TGAAAT

fust-1p::fust-1-ΔN(cDNA)::mRFP (7240 bp)

ΔN cDNA

mRFP

Ampicillin resistance

AAGCttgcatgcctgcaggtcgactctagaggatccccTAGATCGGGCCCTCTTCGTACATTGCGTGGTATGAT
ATCAGCAGGGTATCATCCATTACTCCACATTGAGCCATTTTGTATCGTACCACCTTGTTTGCAT
CTTCTGATGGGATTGGTAAATGACTATATCTTGGCTTCACTTTTGGCACTTGCCAATCAAATCG
ACTTCCCTGGACAGGATCTTCCAGACTCTATGAACGAACAGAGAATGGAATTGGGAATAGTTC
AGAAAGAAGAAAAATTTTGCCAAAAAGTGGTTGAAAGTCTGCAAAACGCCCTTGTGCTCACGA
CGAAAATCAAGGAAGTTTATGAGAGTCGACTAAAGCAGAATAATATCAAACGTTCGTCGATATC
TTGTGGATCTCCAACATGCTGCGTCAGTGATATGCCAAAAAGCCAACAGGATCAAGATATCTT
CTTCTGCGTTGCTTGTCCATTTGCAATTCATATGGTGTGTGCTGGTAGATATACGGAGGATCAA
CGCCAGGCAGCTAAACATCATCCGACTGCTTGTTTAAAATGTCAAGAAAAACATATTCTTTCTA
ATGAGGAAATGATTGAACATGCTAATGAGGCACTTGATGAGCTTGCCAGGCGCTTACAACACG
AGACAAGCCATCTTAGTAACTTGAAAGCTGCTCGCCAAGACATGACAACGGCTATTTCGACTC
GAACTGGGCCAACAAAACAGAAACTCGAGGAAGTTTTGGATTATATTGGCTGTAGTCAAAAAA
TTAACTATCAGGGATTAACCGGAAATCAAGTAAATTTGTGATTTTCGGTATTTGAAAATATTCTA
ATCTTTAGGTTCGCCAACTACTAAGGACAGAAAACATTGACAAAGTACTAAAAGTCTTCACGGA
CTGTAGTTGGATCCAGAATATGAGAAACTTCATGAATGATTTGGCGTCACTCATGTCAGCAAG
CAATAACGCAGTCTTCACAGATGATGATTTGAATGATTTCGCTAAAACTTTGAAGAACTTGAAG
AGAAACATCAGAATTCTTCATCCTTCGATGGGAGTCACTCCTAAACTTCACATCCTCTGTACTC
ACCTCGAGCCTTACATTAGAGCAAAAAGAACTTGGGGTCGTACTTCAGAACAAGGGATGGAAG
CGTTTCATGTGCTCTTCAGGACAACAGAGACGCGATTTTCTTCTGTGATGTCTCTCAAGCTAC
GCGCTGAGTTGATGATCGACTATTTCGCTAACATCAACTATATTTCCGATAAAAGAAACTAGTG
ACAGCAAACATTTCTAGGAAACCGTTTTTATAGTACAtctctctctctctctcactcACATGttttttttCGATTA
AATCAATTAAACGAGCAGAAAAGCCATGAATGATCTCTAAATTCCGACTTACAAAGTGTTTTGA
GTTGATGATAAGCGTGAATCCTAATTTTCAATTGCCTAAACTAACAAACGGGTTGTGAATGGGT
TCAGATTGGGGTACAGCATGAAAATatttcaaaagaagtgaatacaattttagaaaaaccgaatacaaGCAATGTA
AACATGaaaatgaaacaaaaaaacgaaaaaaaaaTCAACCTTGTTCGAAGAAAATGATGAGAATTAGAAT
GAGCAAGGAAATGCGAGAGACTTGGAAATGTGAGAGACGCAGCATTTCACGCGAAACGCGCC
AGCAATCGTCGCGCTTGCGCATTTCGAGACGAGGATACTGTGTAAAATTGTAATCATTTGTTG
CAATTTTTGTTCTTGAGTTAACTTTATTTTTCATTTTCCTCGACCTAATTTCTCAAAATTGTTAAA
ATGTATACTATTTATTACCCTGCAGTGCTCAAATCTTGTGATGAGCACTATCAAAGTTTTAGATG
GAGttttttaaaaagtatttttttttagtataattgttttttttcagtaaatttagtttttttGGAGCAGAACGCTGAAATCTGCCAAGGT
TCGCTACAACCAGATATATCTAGATATAACTGCAAATGTGCGCGCTCCATTGAGACTACACGC
GACGCACAACCGCGACGCGTATTGGACAGCGTTCTCCGTCTTCTTTGAACAGCATCTGTCCTC
AACCATTCTCTTCTCGTTTCTCTTTTAATTTTAAGATATATTTGCTCTTCAAAGGATAAATTTCAG
TCAACGCATCGACACGTAGCGAGTGCCTCGACTTGAATATCGACAAAAATGCGAGTTGAGTTG
AAGGAGACCGTCTTTGTGCAGGGAATTTCCACCACTGCAAATGAAGCCTACATCGCCGACGT
CTTCAGTACCTGCGGAGATATTGCAAAGAACGATCGCGGACCGAGAATCAAGATTTACACCGA
TCGCAACACCGGAGAACCAAAAGGAGAATGCATGATAACATTCGTCGATGCTTCTGCAGCTCA
ACAAGCTATAACTATGTACAATGGGCAACCATTCCCTGGCGGCTCAAGCCCGATGAGTATTTC
ACTAGCCAAGTTCCGCGCTGATGCAGGAGGTGAACGAGGTGGTCGTGGTGGACGAGGCGGT
TTCGGAGGTGGTCGTGGTGGTCCTATGGGAGGACGTGGTGGCTTTGgcggtgatcgtggaggatatggc
ggcggcggtggtcgtggaggattcgatggtggtcgcggtggtggcggtggattccgtggcggagaccgaggaggattccgaggtggtga
tagaggaggcttccgcggaggagatcgtggaggattccgaggcggcgaccgtggtggcgatcgtggaggattcagaggaggCCGT
GGAGTAGGTGGCGGAAATGCTAATATGGAACAAAGGAAGAATGACTGGCCATGTGAGCAGTG
CGGAAATAGCAACTTCGCTTTCAGAAGAGAATGCAATCAATGCCAAGCCCCGAGACCAGATG
GAGGATCCGGAGGTGGTGGTGGCGAGCGACGAGGAGGACCACCAGGAGGTGACCGATACC
GTCCATATgggattggCCAAAGGACCCAAAGGTATGTTTCGAATGATACTAACATAACATAGAACA
TTTTCAGGAGGACCCTTGCTTGGAGGGTACCGGTAGAAAAAATGGCCTCCTCCGAGGACGTC
ATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCGTGAACGGCCACGAGTTCGA
GATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGT
GACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCCAGTACGGCT
CCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAG
GGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGG
ACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCC

TCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCACCGAGCGGATGT
ACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGATGAGGCTGAAGCTGAAGGACGGCG
GCCACTACGACGCCGAGGTCAAGACCACCTACATGGCCAAGAAGCCCGTGCAGCTGCCCGG
CGCCTACAAGACCGACATCAAGCTGGACATCACCTCCCACAACGAGGACTACACCATCGTGG
AACAGTACGAGCGCGCCGAGGGCCGCCACTCCACCGGCGCCTAAGAATTCCAACTGAGCGC
CGGTCGCTACCATTACCAACTTGTCTGGTGTCAAAAATAATAGGGGCCGCTGTCATCAGAGTA
AGTTTAAACTGAGTTCTACTAACTAACGAGTAATATTTAAATTTTCAGCTCTCGCGCCCGTGCC
TCTGACTTCTAAGTCCAATTACTCTTCAACATCCCTACATGCTCTTTCTCCCTGTGCTCCCACC
CCCTATTTTTGTTATTATCAAAAAAACTTCTTCTTAATTTCTTTGTTTTTTAGCTTCTTTTAAGTCA
CCTCTAACAATGAAATTGTGTAGATTCAAAAATAGAATTAATTCGTAATAAAAAGTCGAAAAAAA
TTGTGCTCCCTCCCCCCATTAATAATAATTCTATCCCAAAATCTACACAATGTTCTGTGTACACT
TCTTATGTTTTTTTTACTTCTGATAAATTTTTTTTGAAACATCATAGAAAAAACCGCACACAAAAT
ACCTTATCATATGTTACGTTTCAGTTTATGACCGCAATTTTTATTTCTTCGCACGTCTGGGCCTC
TCATGACGTCAAATCATGCTCATCGTGAAAAAGTTTTGGAGTATTTTTGGAATTTTTCAATCAAG
TGAAAGTTTATGAAATTAATTTTCCTGCTTTTGCTTTTTGGGGGTTTCCCCTATTGTTTGTCAAG
AGTTTCGAGGACGGCGTTTTTCTTGCTAAAATCACAAGTATTGATGAGCACGATGCAAGAAAG
ATCGGAAGAAGGTTTGGGTTTGAGGCTCAGTGGAAGGTGAGTAGAAGTTGATAATTTGAAAGT
GGAGTAGTGTCTATGGGGTTTTTGCCTTAAATGACAGAATACATTCCCAATATACCAAACATAA
CTGTTTCCTACTAGTCGGCCGTACGGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGA
AAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGA
GCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTA
TGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGAT
GCGTAAGGAGAAAATACCGCATCAGGCGGCCTTAAGGGCCTCGTGATACGCCTATTTTTATAG
GTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGC
GGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACC
CTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCC
CTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAG
TAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGC
GGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTC
TGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATA
CACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGC
ATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTA
CTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCA
TGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTG
ACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTA
CTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTC
TGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGG
TCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTAC
ACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTC
ACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAAC
TTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCT
TAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGA
GATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTG
GTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCG
CAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTA
GCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAA
GTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCT
GAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATA
CCTACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTAT
CCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCC
TGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCT
CGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGC
CTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGT
ATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGT
CAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCC
GATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACG
CAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCG
TATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTAC
GCCAAGCTgtaagtttaaacatgatcttactaactaactattctcatttaaattttcagAGCTTAAAAATGGCTGAAATCAC
TCACAACGATGGATACGCTAACAACTTGGAAATGAAAT

fust-1p::fust-1a(cDNA)::unc-54 3'UTR (7108 bp)

isoform a cDNA

Ampicillin resistance

AAGCttgcatgcctgcaggtcgactctagaggatccccTAGATCGGGCCCTCTTCGTACATTGCGTGGTATGAT
ATCAGCAGGGTATCATCCATTACTCCACATTGAGCCATTTTGTATCGTACCACCTTGTTTGCAT
CTTCTGATGGGATTGGTAAATGACTATATCTTGGCTTCACTTTTGGCACTTGCCAATCAAATCG
ACTTCCCTGGACAGGATCTTCCAGACTCTATGAACGAACAGAGAATGGAATTGGGAATAGTTC
AGAAAGAAGAAAAATTTTGCCAAAAAGTGGTTGAAAGTCTGCAAAACGCCCTTGTGCTCACGA
CGAAAATCAAGGAAGTTTATGAGAGTCGACTAAAGCAGAATAATATCAAACGTTCGTCGATATC
TTGTGGATCTCCAACATGCTGCGTCAGTGATATGCCAAAAAGCCAACAGGATCAAGATATCTT
CTTCTGCGTTGCTTGTCCATTTGCAATTCATATGGTGTGTGCTGGTAGATATACGGAGGATCAA
CGCCAGGCAGCTAAACATCATCCGACTGCTTGTTTAAAATGTCAAGAAAAACATATTCTTTCTA
ATGAGGAAATGATTGAACATGCTAATGAGGCACTTGATGAGCTTGCCAGGCGCTTACAACACG
AGACAAGCCATCTTAGTAACTTGAAAGCTGCTCGCCAAGACATGACAACGGCTATTTCGACTC
GAACTGGGCCAACAAAACAGAAACTCGAGGAAGTTTTGGATTATATTGGCTGTAGTCAAAAAA
TTAACTATCAGGGATTAACCGGAAATCAAGTAAATTTGTGATTTTCGGTATTTGAAAATATTCTA
ATCTTTAGGTTCGCCAACTACTAAGGACAGAAAACATTGACAAAGTACTAAAAGTCTTCACGGA
CTGTAGTTGGATCCAGAATATGAGAAACTTCATGAATGATTTGGCGTCACTCATGTCAGCAAG
CAATAACGCAGTCTTCACAGATGATGATTTGAATGATTTCGCTAAAACTTTGAAGAACTTGAAG
AGAAACATCAGAATTCTTCATCCTTCGATGGGAGTCACTCCTAAACTTCACATCCTCTGTACTC
ACCTCGAGCCTTACATTAGAGCAAAAAGAACTTGGGGTCGTACTTCAGAACAAGGGATGGAAG
CGTTTCATGTGCTCTTCAGGACAACAGAGACGCGATTTTCTTCTGTGATGTCTCTCAAGCTAC
GCGCTGAGTTGATGATCGACTATTTCGCTAACATCAACTATATTTCCGATAAAAGAAACTAGTG
ACAGCAAACATTTCTAGGAAACCGTTTTTATAGTACAtctctctctctctctctcactcACATGttttttttCGATTA
AATCAATTAAACGAGCAGAAAAGCCATGAATGATCTCTAAATTCCGACTTACAAAGTGTTTTGA
GTTGATGATAAGCGTGAATCCTAATTTTCAATTGCCTAAACTAACAAACGGGTTGTGAATGGGT
TCAGATTGGGGTACAGCATGAAAATatttcaaaagaagtgaatacaattttagaaaaaccgaatacaaGCAATGTA
AACATGaaaatgaaacaaaaaaacgaaaaaaaaaTCAACCTTGTTCGAAGAAAATGATGAGAATTAGAAT
GAGCAAGGAAATGCGAGAGACTTGGAAATGTGAGAGACGCAGCATTTCACGCGAAACGCGCC
AGCAATCGTCGCGCTTGCGCATTTCGAGACGAGGATACTGTGTAAAATTGTAATCATTTGTTG
CAATTTTTGTTCTTGAGTTAACTTTATTTTTCATTTTCCTCGACCTAATTTCTCAAAATTGTTAAA
ATGTATACTATTTATTACCCTGCAGTGCTCAAATCTTGTGATGAGCACTATCAAAGTTTTAGATG
GAGttttttaaaaagtatttttttttagtataattgttttttttcagtaaatttagtttttttGGAGCAGAACGCTGAAATCTGCCAAGGT
TCGCTACAACCAGATATATCTAGATATAACTGCAAATGTGCGCGCTCCATTGAGACTACACGC
GACGCACAACCGCGACGCGTATTGGACAGCGTTCTCCGTCTTCTTTGAACAGCATCTGTCCTC
AACCATTCTCTTCTCGTTTCTCTTTTAATTTTAAGATATATTTGCTCTTCAAAGGATAAATTTCAG
TCAACGCATCGACACGTAGCGAGTGCCTCGACTTGAATATCGACAAAAATGGCGGCTTACGA
CCAAAGTCAACCCGACTACAGCACACCCGAAGGGCAGCAGGCGTATTGGGCCTATTATcaaca
gcaacaacaacaacaacCAGGAGGTCAACCTGATCAGGATCCGTACGCCGCTGCTGCTTATGGAGG
CCACGACCAAGCACAACAACCGCAAAATCCATATGCACCACCACCACCAGGAGCGGATCCAT
ACGGCCAAGGATCCGGAGGCCAATCCGGAGGATCTGACCCCTACGGGCAAAGTAGAGGTGG
CGGCCGTGGAGGATTCGGAGGCAGTCGTGGAGGAGGTGGATATGATGGTGGACGTGGCGG
AAGTCGTggaggatacgacggaggacgcggaggttacggcggtgatcgcggaggaCGTGGTGGTGGTAGAGGTG
GCTACGACGGAGAACGCCGAGGTGGGAGCCGATGGGATGACGGAAACTCTGATCGACAGGG
CGGACCTCCAGGAGGAAGAGGAGGTTATCAAGACAGAGGACCCCGACGTGACGGACCACCA
AGTGGAGGAGGATACGGAGGCGGTGGTGCTGCTTCAGGAAACCGCGAATTCGGATCCGATG
GGCGAGTTGAGTTGAAGGAGACCGTCTTTGTGCAGGGAATTTCCACCACTGCAAATGAAGCC
TACATCGCCGACGTCTTCAGTACCTGCGGAGATATTGCAAAGAACGATCGCGGACCGAGAAT
CAAGATTTACACCGATCGCAACACCGGAGAACCAAAAGGAGAATGCATGATAACATTCGTCGA
TGCTTCTGCAGCTCAACAAGCTATAACTATGTACAATGGGCAACCATTCCCTGGCGGCTCAAG
CCCGATGAGTATTTCACTAGCCAAGTTCCGCGCTGATGCAGGAGGTGAACGAGGTGGTCGTG
GTGGACGAGGCGGTTTCGGAGGTGGTCGTGGTGGTCCTATGGGAGGACGTGGTGGCTTTGg
cggtgatcgtggaggatatggcggcggcggtggtcgtggaggattcgatggtggtcgcggtggtggcggtggattccgtggcggagaccg
aggaggattccgaggtggtgatagaggaggcttccgcggaggagatcgtggaggattccgaggcggcgaccgtggtggcgatcgtgga
ggattcagaggaggCCGTGGAGTAGGTGGCGGAAATGCTAATATGGAACAAAGGAAGAATGACTGG
CCATGTGAGCAGTGCGGAAATAGCAACTTCGCTTTCAGAAGAGAATGCAATCAATGCCAAGCC
CCGAGACCAGATGGAGGATCCGGAGGTGGTGGTGGCGAGCGACGAGGAGGACCACCAGGA
GGTGACCGATACCGTCCATATGAATTCCAACTGAGCGCCGGTCGCTACCATTACCAACTTGTC

TGGTGTCAAAAATAATAGGGGCCGCTGTCATCAGAGTAAGTTTAAACTGAGTTCTACTAACTAA
CGAGTAATATTTAAATTTTCAGCTCTCGCGCCCGTGCCTCTGACTTCTAAGTCCAATTACTCTT
CAACATCCCTACATGCTCTTTCTCCCTGTGCTCCCACCCCCTATTTTTGTTATTATCAAAAAAAC
TTCTTCTTAATTTCTTTGTTTTTTAGCTTCTTTTAAGTCACCTCTAACAATGAAATTGTGTAGATT
CAAAAATAGAATTAATTCGTAATAAAAAGTCGAAAAAAATTGTGCTCCCTCCCCCCATTAATAAT
AATTCTATCCCAAAATCTACACAATGTTCTGTGTACACTTCTTATGTTTTTTTTACTTCTGATAAA
TTTTTTTTGAAACATCATAGAAAAAACCGCACACAAAATACCTTATCATATGTTACGTTTCAGTT
TATGACCGCAATTTTTATTTCTTCGCACGTCTGGGCCTCTCATGACGTCAAATCATGCTCATCG
TGAAAAAGTTTTGGAGTATTTTTGGAATTTTTCAATCAAGTGAAAGTTTATGAAATTAATTTTCCT
GCTTTTGCTTTTTTGGGGGTTTCCCCTATTGTTTGTCAAGAGTTTCGAGGACGGCGTTTTTCTTG
CTAAAATCACAAGTATTGATGAGCACGATGCAAGAAAGATCGGAAGAAGGTTTGGGTTTGAGG
CTCAGTGGAAGGTGAGTAGAAGTTGATAATTTGAAAGTGGAGTAGTGTCTATGGGGTTTTTGC
CTTAAATGACAGAATACATTCCCAATATACCAAACATAACTGTTTCCTACTAGTCGGCCGTACG
GGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCG
GAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGT
CAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACT
GAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCA
GGCGGCCTTAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTT
CTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTA
AATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAA
AAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTG
CCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGG
TGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCC
CGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGT
ATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGA
GTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGC
TGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAA
GGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACC
GGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAA
CAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGA
CTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGG
TTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGG
GCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGG
ATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAG
ACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGG
TGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGC
GTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGC
TGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCA
ACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTG
TAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTA
ATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAG
ACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCC
AGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAAGCGC
CACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGA
GAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCG
CCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAA
ACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCT
TTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCG
CTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCC
CAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGG
TTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAG
GCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAA
CAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTgtaagtttaaacatgatcttactaactaac
tattctcatttaaattttcagAGCTTAAAAATGGCTGAAATCACTCACAACGATGGATACGCTAACAACTTG
GAAATGAAAT

fust-1p::fust-1b(cDNA)::unc-54 3'UTR (6874 bp)

isoform b cDNA

Ampicillin resistance

AAGCttgcatgcctgcaggtcgactctagaggatccccTAGATCGGGCCCTCTTCGTACATTGCGTGGTATGAT
ATCAGCAGGGTATCATCCATTACTCCACATTGAGCCATTTTGTATCGTACCACCTTGTTTGCAT
CTTCTGATGGGATTGGTAAATGACTATATCTTGGCTTCACTTTTGGCACTTGCCAATCAAATCG
ACTTCCCTGGACAGGATCTTCCAGACTCTATGAACGAACAGAGAATGGAATTGGGAATAGTTC
AGAAAGAAGAAAAATTTTGCCAAAAAGTGGTTGAAAGTCTGCAAAACGCCCTTGTGCTCACGA
CGAAAATCAAGGAAGTTTATGAGAGTCGACTAAAGCAGAATAATATCAAACGTTCGTCGATATC
TTGTGGATCTCCAACATGCTGCGTCAGTGATATGCCAAAAAGCCAACAGGATCAAGATATCTT
CTTCTGCGTTGCTTGTCCATTTGCAATTCATATGGTGTGTGCTGGTAGATATACGGAGGATCAA
CGCCAGGCAGCTAAACATCATCCGACTGCTTGTTTAAAATGTCAAGAAAAACATATTCTTTCTA
ATGAGGAAATGATTGAACATGCTAATGAGGCACTTGATGAGCTTGCCAGGCGCTTACAACACG
AGACAAGCCATCTTAGTAACTTGAAAGCTGCTCGCCAAGACATGACAACGGCTATTTCGACTC
GAACTGGGCCAACAAAACAGAAACTCGAGGAAGTTTTGGATTATATTGGCTGTAGTCAAAAAA
TTAACTATCAGGGATTAACCGGAAATCAAGTAAATTTGTGATTTTCGGTATTTGAAAATATTCTA
ATCTTTAGGTTCGCCAACTACTAAGGACAGAAAACATTGACAAAGTACTAAAAGTCTTCACGGA
CTGTAGTTGGATCCAGAATATGAGAAACTTCATGAATGATTTGGCGTCACTCATGTCAGCAAG
CAATAACGCAGTCTTCACAGATGATGATTTGAATGATTTCGCTAAAACTTTGAAGAACTTGAAG
AGAAACATCAGAATTCTTCATCCTTCGATGGGAGTCACTCCTAAACTTCACATCCTCTGTACTC
ACCTCGAGCCTTACATTAGAGCAAAAAGAACTTGGGGTCGTACTTCAGAACAAGGGATGGAAG
CGTTTCATGTGCTCTTCAGGACAACAGAGACGCGATTTTCTTCTGTGATGTCTCTCAAGCTAC
GCGCTGAGTTGATGATCGACTATTTCGCTAACATCAACTATATTTCCGATAAAAGAAACTAGTG
ACAGCAAACATTTCTAGGAAACCGTTTTTATAGTACAtctctctctctctctctcactcACATGttttttttCGATTA
AATCAATTAAACGAGCAGAAAAGCCATGAATGATCTCTAAATTCCGACTTACAAAGTGTTTTGA
GTTGATGATAAGCGTGAATCCTAATTTTCAATTGCCTAAACTAACAAACGGGTTGTGAATGGGT
TCAGATTGGGGTACAGCATGAAAATatttcaaaagaagtgaatacaattttagaaaaaccgaatacaaGCAATGTA
AACATGaaaatgaaacaaaaaaacgaaaaaaaaaTCAACCTTGTTCGAAGAAAATGATGAGAATTAGAAT
GAGCAAGGAAATGCGAGAGACTTGGAAATGTGAGAGACGCAGCATTTCACGCGAAACGCGCC
AGCAATCGTCGCGCTTGCGCATTTCGAGACGAGGATACTGTGTAAAATTGTAATCATTTGTTG
CAATTTTTGTTCTTGAGTTAACTTTATTTTTCATTTTCCTCGACCTAATTTCTCAAAATTGTTAAA
ATGTATACTATTTATTACCCTGCAGTGCTCAAATCTTGTGATGAGCACTATCAAAGTTTTAGATG
GAGttttttaaaaagtatttttttttagtataattgttttttttcagtaaatttagtttttttGGAGCAGAACGCTGAAATCTGCCAAGGT
TCGCTACAACCAGATATATCTAGATATAACTGCAAATGTGCGCGCTCCATTGAGACTACACGC
GACGCACAACCGCGACGCGTATTGGACAGCGTTCTCCGTCTTCTTTGAACAGCATCTGTCCTC
AACCATTCTCTTCTCGTTTCTCTTTTAATTTTAAGATATATTTGCTCTTCAAAGGATAAATTTCAG
TCAACGCATCGACACGTAGCGAGTGCCTCGACTTGAATATCGACAAAAATGGAGGCCACGAC
CAAGCACAACAACCGCAAAATCCATATGCACCACCACCACCAGGAGCGGATCCATACGGCCA
AGGATCCGGAGGCCAATCCGGAGGATCTGACCCCTACGGGCAAAGTAGAGGTGGCGGCCGT
GGAGGATTCGGAGGCAGTCGTGGAGGAGGTGGATATGATGGTGGACGTGGCGGAAGTCGTg
gaggatacgacggaggacgcggaggttacggcggtgatcgcggaggaCGTGGTGGTGGTAGAGGTGGCTACGAC
GGAGAACGCCGAGGTGGGAGCCGATGGGATGACGGAAACTCTGATCGACAGAGGACCCCGA
CGTGACGGACCACCAAGTGGAGGAGGATACGGAGGCGGTGGTGCTGCTTCAGGAAACCGCG
AATTCGGATCCGATGGGCGAGTTGAGTTGAAGGAGACCGTCTTTGTGCAGGGAATTTCCACC
ACTGCAAATGAAGCCTACATCGCCGACGTCTTCAGTACCTGCGGAGATATTGCAAAGAACGAT
CGCGGACCGAGAATCAAGATTTACACCGATCGCAACACCGGAGAACCAAAAGGAGAATGCAT
GATAACATTCGTCGATGCTTCTGCAGCTCAACAAGCTATAACTATGTACAATGGGCAACCATTC
CCTGGCGGCTCAAGCCCGATGAGTATTTCACTAGCCAAGTTCCGCGCTGATGCAGGAGGTGA
ACGAGGTGGTCGTGGTGGACGAGGCGGTTTCGGAGGTGGTCGTGGTGGTCCTATGGGAGGA
CGTGGTGGCTTTGgcggtgatcgtggaggatatggcggcggcggtggtcgtggaggattcgatggtggtcgcggtggtggcggt
ggattccgtggcggagaccgaggaggattccgaggtggtgatagaggaggcttccgcggaggagatcgtggaggattccgaggcggc
gaccgtggtggcgatcgtggaggattcagaggaggCCGTGGAGTAGGTGGCGGAAATGCTAATATGGAACAA
AGGAAGAATGACTGGCCATGTGAGCAGTGCGGAAATAGCAACTTCGCTTTCAGAAGAGAATG
CAATCAATGCCAAGCCCCGAGACCAGATGGAGGATCCGGAGGTGGTGGTGGCGAGCGACGA
GGAGGACCACCAGGAGGTGACCGATACCGTCCATATGAATTCCAACTGAGCGCCGGTCGCTA
CCATTACCAACTTGTCTGGTGTCAAAAATAATAGGGGCCGCTGTCATCAGAGTAAGTTTAAACT
GAGTTCTACTAACTAACGAGTAATATTTAAATTTTCAGCTCTCGCGCCCGTGCCTCTGACTTCT
AAGTCCAATTACTCTTCAACATCCCTACATGCTCTTTCTCCCTGTGCTCCCACCCCCTATTTTT

GTTATTATCAAAAAAACTTCTTCTTAATTTCTTTGTTTTTTAGCTTCTTTTAAGTCACCTCTAACA
ATGAAATTGTGTAGATTCAAAAATAGAATTAATTCGTAATAAAAAGTCGAAAAAAATTGTGCTCC
CTCCCCCCATTAATAATAATTCTATCCCAAAATCTACACAATGTTCTGTGTACACTTCTTATGTT
TTTTTTACTTCTGATAAATTTTTTTTGAAACATCATAGAAAAAACCGCACACAAAATACCTTATCA
TATGTTACGTTTCAGTTTATGACCGCAATTTTTATTTCTTCGCACGTCTGGGCCTCTCATGACG
TCAAATCATGCTCATCGTGAAAAAGTTTTGGAGTATTTTTGGAATTTTTCAATCAAGTGAAAGTT
TATGAAATTAATTTTCCTGCTTTTGCTTTTTGGGGGTTTCCCCTATTGTTTGTCAAGAGTTTCGA
GGACGGCGTTTTTCTTGCTAAAATCACAAGTATTGATGAGCACGATGCAAGAAAGATCGGAAG
AAGGTTTGGGTTTGAGGCTCAGTGGAAGGTGAGTAGAAGTTGATAATTTGAAAGTGGAGTAGT
GTCTATGGGGTTTTTGCCTTAAATGACAGAATACATTCCCAATATACCAAACATAACTGTTTCCT
ACTAGTCGGCCGTACGGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTG
ACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAA
GCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCAT
CAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGG
AGAAAATACCGCATCAGGCGGCCTTAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGT
CATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCC
TATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAAT
GCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCC
TTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATG
CTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATC
CTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTG
GCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCT
CAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTA
AGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACA
ACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCG
CCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGA
TGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTT
CCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCG
GCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGG
TATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGG
GGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTA
AGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTT
TAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGA
GTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTT
TTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTG
CCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCA
AATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCT
ACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTT
ACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGG
GTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGT
GAGCATTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCG
GCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTA
TAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGG
GCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGC
CTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTT
TGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAG
GAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATG
CAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGA
GTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTG
GAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTgta
agtttaaacatgatcttactaactaactattctcatttaaattttcagAGCTTAAAAATGGCTGAAATCACTCACAACGAT
GGATACGCTAACAACTTGGAAATGAAAT

fust-1p::fust-1-ΔN(cDNA)::unc-54 3'UTR (6466 bp)

ΔN cDNA

Ampicillin resistance

AAGCttgcatgcctgcaggtcgactctagaggatccccTAGATCGGGCCCTCTTCGTACATTGCGTGGTATGAT
ATCAGCAGGGTATCATCCATTACTCCACATTGAGCCATTTTGTATCGTACCACCTTGTTTGCAT
CTTCTGATGGGATTGGTAAATGACTATATCTTGGCTTCACTTTTGGCACTTGCCAATCAAATCG
ACTTCCCTGGACAGGATCTTCCAGACTCTATGAACGAACAGAGAATGGAATTGGGAATAGTTC
AGAAAGAAGAAAAATTTTGCCAAAAAGTGGTTGAAAGTCTGCAAAACGCCCTTGTGCTCACGA
CGAAAATCAAGGAAGTTTATGAGAGTCGACTAAAGCAGAATAATATCAAACGTTCGTCGATATC
TTGTGGATCTCCAACATGCTGCGTCAGTGATATGCCAAAAAGCCAACAGGATCAAGATATCTT
CTTCTGCGTTGCTTGTCCATTTGCAATTCATATGGTGTGTGCTGGTAGATATACGGAGGATCAA
CGCCAGGCAGCTAAACATCATCCGACTGCTTGTTTAAAATGTCAAGAAAAACATATTCTTTCTA
ATGAGGAAATGATTGAACATGCTAATGAGGCACTTGATGAGCTTGCCAGGCGCTTACAACACG
AGACAAGCCATCTTAGTAACTTGAAAGCTGCTCGCCAAGACATGACAACGGCTATTTCGACTC
GAACTGGGCCAACAAAACAGAAACTCGAGGAAGTTTTGGATTATATTGGCTGTAGTCAAAAAA
TTAACTATCAGGGATTAACCGGAAATCAAGTAAATTTGTGATTTTCGGTATTTGAAAATATTCTA
ATCTTTAGGTTCGCCAACTACTAAGGACAGAAAACATTGACAAAGTACTAAAAGTCTTCACGGA
CTGTAGTTGGATCCAGAATATGAGAAACTTCATGAATGATTTGGCGTCACTCATGTCAGCAAG
CAATAACGCAGTCTTCACAGATGATGATTTGAATGATTTCGCTAAAACTTTGAAGAACTTGAAG
AGAAACATCAGAATTCTTCATCCTTCGATGGGAGTCACTCCTAAACTTCACATCCTCTGTACTC
ACCTCGAGCCTTACATTAGAGCAAAAAGAACTTGGGGTCGTACTTCAGAACAAGGGATGGAAG
CGTTTCATGTGCTCTTCAGGACAACAGAGACGCGATTTTCTTCTGTGATGTCTCTCAAGCTAC
GCGCTGAGTTGATGATCGACTATTTCGCTAACATCAACTATATTTCCGATAAAAGAAACTAGTG
ACAGCAAACATTTCTAGGAAACCGTTTTTATAGTACAtctctctctctctctctcactcACATGttttttttCGATTA
AATCAATTAAACGAGCAGAAAAGCCATGAATGATCTCTAAATTCCGACTTACAAAGTGTTTTGA
GTTGATGATAAGCGTGAATCCTAATTTTCAATTGCCTAAACTAACAAACGGGTTGTGAATGGGT
TCAGATTGGGGTACAGCATGAAAATatttcaaaagaagtgaatacaattttagaaaaaccgaatacaaGCAATGTA
AACATGaaaatgaaacaaaaaaacgaaaaaaaaaTCAACCTTGTTCGAAGAAAATGATGAGAATTAGAAT
GAGCAAGGAAATGCGAGAGACTTGGAAATGTGAGAGACGCAGCATTTCACGCGAAACGCGCC
AGCAATCGTCGCGCTTGCGCATTTCGAGACGAGGATACTGTGTAAAATTGTAATCATTTGTTG
CAATTTTTGTTCTTGAGTTAACTTTATTTTTCATTTTCCTCGACCTAATTTCTCAAAATTGTTAAA
ATGTATACTATTTATTACCCTGCAGTGCTCAAATCTTGTGATGAGCACTATCAAAGTTTTAGATG
GAGttttttaaaaagtatttttttttagtataattgttttttcagtaaatttagtttttGGAGCAGAACGCTGAAATCTGCCAAGGT
TCGCTACAACCAGATATATCTAGATATAACTGCAAATGTGCGCGCTCCATTGAGACTACACGC
GACGCACAACCGCGACGCGTATTGGACAGCGTTCTCCGTCTTCTTTGAACAGCATCTGTCCTC
AACCATTCTCTTCTCGTTTCTCTTTTAATTTTAAGATATATTTGCTCTTCAAAGGATAAATTTCAG
TCAACGCATCGACACGTAGCGAGTGCCTCGACTTGAATATCGACAAAAATGCGAGTTGAGTTG
AAGGAGACCGTCTTTGTGCAGGGAATTTCCACCACTGCAAATGAAGCCTACATCGCCGACGT
CTTCAGTACCTGCGGAGATATTGCAAAGAACGATCGCGGACCGAGAATCAAGATTTACACCGA
TCGCAACACCGGAGAACCAAAAGGAGAATGCATGATAACATTCGTCGATGCTTCTGCAGCTCA
ACAAGCTATAACTATGTACAATGGGCAACCATTCCCTGGCGGCTCAAGCCCGATGAGTATTTC
ACTAGCCAAGTTCCGCGCTGATGCAGGAGGTGAACGAGGTGGTCGTGGTGGACGAGGCGGT
TTCGGAGGTGGTCGTGGTGGTCCTATGGGAGGACGTGGTGGCTTTGgcggtgatcgtggaggatatggc
ggcggcggtggtcgtggaggattcgatggtggtcgcggtggtggcggtggattccgtggcggagaccgaggaggattccgaggtggtga
tagaggaggcttccgcggaggagatcgtggaggattccgaggcggcgaccgtggtggcgatcgtggaggattcagaggaggCCGT
GGAGTAGGTGGCGGAAATGCTAATATGGAACAAAGGAAGAATGACTGGCCATGTGAGCAGTG
CGGAAATAGCAACTTCGCTTTCAGAAGAGAATGCAATCAATGCCAAGCCCCGAGACCAGATG
GAGGATCCGGAGGTGGTGGTGGCGAGCGACGAGGAGGACCACCAGGAGGTGACCGATACC
GTCCATATGAATTCCAACTGAGCGCCGGTCGCTACCATTACCAACTTGTCTGGTGTCAAAAAT
AATAGGGGCCGCTGTCATCAGAGTAAGTTTAAACTGAGTTCTACTAACTAACGAGTAATATTTA
AATTTTCAGCTCTCGCGCCCGTGCCTCTGACTTCTAAGTCCAATTACTCTTCAACATCCCTACA
TGCTCTTTCTCCCTGTGCTCCCACCCCCTATTTTTGTTATTATCAAAAAAACTTCTTCTTAATTTC
TTTGTTTTTTAGCTTCTTTTAAGTCACCTCTAACAATGAAATTGTGTAGATTCAAAAATAGAATTA
ATTCGTAATAAAAAGTCGAAAAAAATTGTGCTCCCTCCCCCCATTAATAATAATTCTATCCCAAA
ATCTACACAATGTTCTGTGTACACTTCTTATGTTTTTTTTACTTCTGATAAATTTTTTTTGAAACA
TCATAGAAAAAACCGCACACAAAATACCTTATCATATGTTACGTTTCAGTTTATGACCGCAATTT
TTATTTCTTCGCACGTCTGGGCCTCTCATGACGTCAAATCATGCTCATCGTGAAAAAGTTTTGG
AGTATTTTTGGAATTTTTCAATCAAGTGAAAGTTTATGAAATTAATTTTCCTGCTTTTGCTTTTTG

GGGGTTTCCCCTATTGTTTGTCAAGAGTTTCGAGGACGGCGTTTTTCTTGCTAAAATCACAAGT
ATTGATGAGCACGATGCAAGAAAGATCGGAAGAAGGTTTGGGTTTGAGGCTCAGTGGAAGGT
GAGTAGAAGTTGATAATTTGAAAGTGGAGTAGTGTCTATGGGGTTTTTGCCTTAAATGACAGAA
TACATTCCCAATATACCAAACATAACTGTTTCCTACTAGTCGGCCGTACGGGCCCTTTCGTCTC
GCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGC
TTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGC
GGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATAT
GCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGGCCTTAAGGGC
CTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTG
GCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATAT
GTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATG
AGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGC
TCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTT
ACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTC
CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGC
AAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCA
CAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGA
GTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCT
TTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAA
GCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAA
ACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGC
GGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAA
ATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAG
CCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGA
CAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCA
TATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTT
GATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTA
GAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAA
AAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGA
AGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAG
GCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAG
TGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCG
GATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAA
CGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCCGAA
GGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGG
GAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTT
GAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGC
GGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCC
CCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCG
AACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCG
CCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAA
AGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTT
TACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAG
GAAACAGCTATGACCATGATTACGCCAAGCTgtaagtttaaacatgatcttactaactaactattctcatttaaattttcag
AGCTTAAAAATGGCTGAAATCACTCACAACGATGGATACGCTAACAACTTGGAAATGAAAT